

```

QY 862 VKSCTEAEBCP--AKGDAEAATAGALKAEKEGSGSRAPTTAKSGAPODSDSA 919
Db 261 GKS--GEGGLPGOGARFPCTPGIPGVKHGRGYPGLGAKGEA-----GAP----- 305
QY 920 TCSADEVDEAGCDKNRLSPRPSILTPGDPANASPOKPLDKOLKORAAAPPIQVT 979
Db 306 -----GVKGE-----SSGSGENGSP--GPMGRGRLGEGGRGPGPAA 341
QY 980 KVHEPPREDAAPTKAPAPPAPPPONLOPESAPQOGSSPRKSSRPAADKEAFAEA 1039
Db 342 GAR---GNDQOPGAPGPGVPAGGPGFPCAPGAKGAGPTGARGP-----EGA 388
QY 1040 OKLGDPPCWTSGLPFPVPPREVIKASPHADPSAFYAPRGHPLPLGHTAPVLRP 1099
Db 389 QGSRGP-----GNPGSPGAGASGNPGTDGIPGAKSGAGAPLGA 430
QY 1100 PTISNPPLISSAKHPSVLERQIGAI SQMSVQLHVPSHAKAPVGVPTWGLPLPMDPK 1159
Db 431 PGFPGP-----RGPFGP--QGATGLOPK 452
QY 1160 KLA---PFSGVKQOLSPRQAGPESLGVPTAOASVLRGTALGSPGSSITKIPSTR 1216
Db 453 GOAGBPAGIAGFKDQ--GPKGETGAPGAGPAGPAGBEEKRG--ARGE--PGAGPIGPGER 509
QY 1217 -VPDSATTYRGSIT--HGTADVLTKGTITRIIGEDSPRL-----DRGR--EDSL 1263
Db 510 GAFNARCFPGDGLGAPKAP-----GERPSGLAGPKGANGPGRGEGGL 556
QY 1264 PKGHVYEGKGHVLVEGMSVTCSEKEDRSSGPHETA-----PKRTYD 1312
Db 557 PGAGGL--TGRPDA--GPGKIVGSPGAPBEDGR--RGPFGOGARQOPVMPGPKGA-- 610
QY 1313 MMEGRVGA-----ISSASIEGLMGR-----AIPPERHSPHLKEQHHRG-SITGIP 1360
Db 611 --NGEPGAKGKGLAGLRLGLPKDGETGACGPSPGAGERGEGAGPSSGFGILP 668
QY 1361 RSVYEAQDYLREAKLKRBTGTPPPP--SRDLTAAYTQALGPKLRAHGLVAT 1417
Db 669 -----GPPGPPGEGKQDQGIPEAGAP-----GLVGP 697
QY 1418 VKEAGRSIHEIPRE-----LRHTPELPLARPLKEGSI-----TQGRP-- 1456
Db 698 RGERG-----PFGERGSGAOLGQPRGLPCTGTDP--KGAAPDPGAGQSPGLOG 750
QY 1457 LKYDTGASTGSKKDVSLIGSPGRTFP-----VHPLDVMAAPALERACEESL 1508
Db 751 MPGERGAGIAGPKGD--RGDVGEKPEGAPGKDGGRGLTGPIGPPGAGA-----NGE 802
QY 1509 KSRGCTASSGGSTARGAPVIVPELGRKPRQSLTYEDHGAPRAGHLPRGSPVTMREPTR 1568
Db 803 KGEVGPSPGSGTGARGAPGETGTP--GP-----AGPAG--PGAD--GQPAK 848
QY 1569 LOEGSISSSKASODKLTSTPREIAKSPHSTVPEHNPPIISYEHILRGVSGVDLYRSHI 1628
Db 849 GDQG-----EAGQ-----KGDAGAPROGSPGAPRGQ-----TGVTPKARG-- 888
QY 1629 PLAFDPTISIPRGIPIDAAAAYVLPRLANPPTYPHLYPPYLRGYPTALLENROTIND 1688
Db 889 --AQP--PGATGPPGAAGRVGPPGANGPG-----PAGPPGAKDGKPGVGRD 934
QY 1689 Y-----ITQOMHNTATMAORA-----DMLRGLSPRESSLALNTAARGIIDL----- 1734
Db 935 SGPPRAGDPGLQGPAGAPGEGEPDGGSLDGPBPQGL--AGQRGIVGLPGQGE 991
QY 1735 SQVPHL--PVLVPPPTGTPATMDRLAYLPTAPQPFSSRHSSSPLSPGAPTHLTKPTTS 1792
Db 992 RGFPLPGSPGPKQAGPAGSGDR-----GPPGVGPPGLT----- 1028
QY 1793 SSERBRDRERDREREKSLITSITVEHAPIRPPTGEOSSGSSGSSGSSSSRPA 1852
Db 1029 -----GPAEPGREGSPGADGPPGRDGA----- 1052

```

```

QY 1853 SHSHAQHSPISPTODALQORPSVLNHTGKGIITAVEBSPVTLNSTSTSSVPAPAT 1912
Db 1053 -----GVKGE-----DRGETGALGAPAP 1070
QY 1913 FPPATHCPLG-----GTLGVYPTLMEVLLPKAPRVARPERPRAD----- 1954
Db 1071 GPPGSPGAPGTGQGDREAGAGQPMQPS-----GPAGARGIAGPQGPBGDKEGSEQ 1124
QY 1955 -----TGHAFLAKPPRASGLRPSASPSKSGSEPRPLVPVPSGHATARTAPAKULA 2003
Db 1125 GERGLKHGFTGLQGLPGRPSPGDQASGPAPGSPGR----- 1163
QY 2004 PHNASPPAPAPASDPHREKTSKPFSTIOELRSLGYHSSSYSPGVPVSPVSSPS 2063
Db 1164 -----GPPGVGPS-----GKDSNGITGPIGPBPGRSG 1194
QY 2064 LTHDKGLPKHLELDKSHLBELRPKOPGVKLGGEAHLPHLRPLPESQSSPPLQTA 2123
Db 1195 ETGVPFGP-----SPGPPGPPGPGIDMSAFAGLGQREKQDPMQY-- 1238
QY 2124 PGVGHQORVVTLAGHISEVITQDYTRH--HPQUSAPLPAPLYFPGASCPVLDLRP-- 2179
Db 1239 --KRADEADSTLRQHDVEV--DATLKSINNQIESIRSPDGSRRNPARTCODLKLCHPEW 1293
QY 2180 PSDLXLPDPDHGAPARQSP--HSEGGKRSPEPNKTSV 2214
Db 1294 KSGDIWIDPNOGCTLDMAKVFCNMEGTGTCVYPAPATV 1331

```

Search completed: April 15, 2004, 22:23:17
Job time : 242 secs

submitted to the EMBL Data Library, August 1990

A:Description: Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B9

A:Reference number: S27923

A:Accession: S27923

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-924 <PAB>

A:Cross-references: EMBL:M35547; NID:g330420; PID:AAA5896.1; PID:g330421

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 2.4% Score 318; DB 2; Length 924;

Best Local Similarity 20.3%; Pred. No. 0.00027;

Matches 284; Conservative 94; Mismatches 453; Indels 566; Gaps 67;

```

QY 731 VPRGCGSPATVNNSSDTEISIPSTEAKDTGNGKPR---PATLGADGPP---GPP 783
DB 4 VARGCCLAP-----GSGGAGHPPRRSGAADPAD---FVGHAPAPRAPGPEPRTLQPA 54
QY 784 TPRRSTRAPITPTASEATGATPPRAP--PSP--SAPPVVYKKEKEEETAAAPVEGE 841
DB 55 TPRRSGAADPADP-----VGHAPAPRAPGPEPRTLQPA---RSGAADPADVPG 103
QY 842 EQKPAAEELAVDTGAKKEPVKSETEAEEGPAGKDAEAATAGALKAKKEGSG 901
DB 104 HPAAPRA-----PGEPRTRLQPA---RSGAADPADVPGHAP--APRAPGPEP 149
QY 902 RA---TTAKSSGAPQSDSSATCSADVDABGDKRLLSPRSLTPTGDPANASP 957
DB 150 RTRLQPA---RSGAADPADPADPVGHAP---PR---APGPEPRTLQ 189
QY 958 QKPELDKQKORAAAIPIQVTKVHEPPREDAPTKAP-----PAPPPONLQPESD 1010
DB 190 ATP-----RSGAADPA-----DVGHPAPRAPGPEPRTLQPA---RSGAADPAD 236
QY 1011 ---APOQCGSSPPKSGSRPAPAPADKEAPAAEAKLPGDPPCWTSGLPFPVPPREVIT 1064
DB 237 PVGHAPAPRAPGPEPRTLQ---PATP---RSGAAD---PADP---VGHAP--- 275
QY 1065 ASBHPADPSAFYAPPGHPLGLHDTAPV--LPRPTISNPPPLISSAKHPSVLERQIG 1123
DB 276 AAPRAPGPEPRTLQPA---RSGAADPADPADPVGHAPAPRAPGPEP 319
QY 1124 AISOQSVOLHNVYSHAKAPVPTMGLPLPMDPKKLAAPFGSVKQEOQLSPR---GQAGCP 1181
DB 320 ---RTRLQPA---RSGAADPADPADVGH---PAAPAPGPEPRTLQPA---RSGAADPAD 371
QY 1182 ESIGVPTAOEASVLRGTALGVGSGITKGIPTSTRVSDSAITYRGSITGTPADVLYKG 1241
DB 372 DVGHPAPRA-----FG-----PGEPRTRLQPA---RSGAAD---ADPADVGH 411
QY 1242 TITRIITGDSPLLDGREGDSLPKGVITYEGKKGHVLSYEGGMSVQCKEDGRSSGCP 1301
DB 412 AAPRAPGPEPRTL----- 425
QY 1302 HETAPAPKRTYDMMEGRVGRASISASIEGLMGRAIPERHSPHILKEOHIRGSIITGIRP 1361
DB 426 -QATPRRS-----GAADPADPVGHAPAPRAPGPEP-----PR 456
QY 1362 SYVEAOEDYLREAKLIREGTPPPPPPSDLTETAYTQALGP---LKLKPAHEGLVAVY 1418
DB 457 T-----RLQPA---RSGAADPADP---VGHAPAPRAPGPEPRTLQ---ATP 498
QY 1419 KEAGRSIHETPRELHTPEPLAPAPRLKSGSIITGTPLYDYGASTGSKKHVDVSLIG 1478
DB 499 RRGAA---DAPDPVGH--PAAPRAPGPEPRTLQPA---RSGA----- 538
QY 1479 SPGETFPVPAPLDVMDARALERAQYEEISLKSRPGTASSGSISARGAPVYELCKPQ 1538
DB 539 -----DAPDPVGHAPRA-----FG-----PE-PRTRL 561
QY 1539 SPLTYEDHGAAPAGHLPGRGSPVTWRETPRLQBOGSSLSASAQDKRLTTPRELAISPHS 1598
DB 562 QPATPRRSAADPAD--FVGHAPAPRAPGPE-----PRTRLQPA---PRR--- 602

```

```

QY 1599 TVPEHHHPISPYEHLNRGVGVVDLYRSHIPLAFDPISPRGIPLDAAAAYLPERHAPN 1658
DB 603 -----SG-----ADPAD--FVGHAP-----APRAGPE 624
QY 1659 PTVPEHLYPYLLINGYPTTALENKQIINDYITSQOMHNHTATAMAQRADMLRGLSPRES 1718
DB 625 P----- 637
QY 1719 SLAINVAAGPRGITDLSQVHLPLVLPVPTGTPATNDRLAVLPTAQPPSSRRSSPLS 1778
DB 638 G-----ADPADPVGHAPAPRAPGPEPRTLQPA---RSGAADPADP---VGHAPAPRA 688
QY 1779 PGFTHLTKPTTSSSERNDREDRDREKSIITSTTVHAPDIWPGTEQSSGS 1838
DB 689 PG-----PEPRTR-----LQPA---RSGAA 708
QY 1839 GSSGGGGSSSRPASHAHQSDISPTODALQORPSVLHNTGMKGIITAVESSKPLVL 1898
DB 709 -----DAPDPVGHAPAPRAPGPEPRTLQPA---LQPA--- 736
QY 1899 RSTSTSSVPRAPATTPPATHTCPLGTLDTGYP--TLMEPVLLPKAPRVARPERRADTGH 1957
DB 737 RSGAADPADVPG--HPAPRAP-----GPEPRTLQPA---TPRSGAADPADVPGH 784
QY 1958 AFLAKPARSGLEPPASSPSKSEPRPLVPVSGHATTARTPAKVLAPHHASP--DPEAP 2015
DB 785 -----PAAPRAPGPEPRTLQPA-----TRRSGAADPADPADPVGHAPAPR 823
QY 2016 ASADPHREKTQKPFSTIQBELRSLGYHSGSYBEGVEPVSPSSPBLTHDKLPHLE 2075
DB 824 APGPEPRTLQPA---RSG---ADPADPVGHAPAPRAPGPEPRTLQ--- 860
QY 2076 ELDKSHLEGELRPKOPG 2092
DB 861 -----WLVKPKGTG 869

```

RESULT 100

B41182

collagen alpha 1(II) chain precursor (long splice form) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1993

C:Accession: B41182

R.Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A:Reference number: B41182; MUID:91358489; PMID:1885613

A:Accession: B41182

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1487 <MET>

A:Cross-references: GB:M65161

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology,

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime

F:33-91/Domain: von Willebrand factor type C repeat homology <WVC>

F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <CC>

Query Match 2.4% Score 318; DB 2; Length 1487;

Best Local Similarity 21.4%; Pred. No. 0.00044;

Matches 354; Conservative 105; Mismatches 619; Indels 580; Gaps 78;

```

QY 721 AEAALHAGNEVRGCGSPATVNNSSDTEISIPSTEAK-----DTGONG--PKP-- 769
DB 90 ADLATASGLGPKGKGGERGDIRDIIGRPGPGPGAGGQGRDRGDKGKGAPGPG 149
QY 770 ---PATLGADP--PGEPTTPRRTS---RAPITPTASEATGA-----PTPP 810
DB 150 RDGEPGTGNPAGPGPGPGGSLAGNFAAQMGAGYDEKAGAGMGVMOGPMGMBR 209
QY 811 APPSPS--APPVVPVKEKEEETAAAPVEGE-----EQKPAAEELAVDTGKAEPP 861
DB 210 GPPGPGAGAP-----QGFQGNFGRGPGVSGPMGRGPGPGAPKPDDEAGKP 260

```


Oy		214	VSPPEPESKERSLVOITTYDENRKKAAEARILLEGLOVEYLUNQPSUT-----RQY	266
Db		206	VIPPPPIASHPT-----PPSHAKTDTGSRHSSGSA	236
Oy		267	HENIKINQAMRKULILEFKSRNARKOMKOCORAYDOLMEALEKKVERIENNPRRAKE	326
Db		237	HSQGFSTPISGSGLIFPDCAKTATHTLYVEI--KDEVIIVEDK-DKETFKPPPNVS-	290
Oy		327	SKVREYEKOFPETIKOREL-----OEMOSRVGORGSLMSAAREHVS	373
Db		291	---NELTAEPERHIRINEMAGIDIEPIOPPIAOERRKSIV---VSGLNMFVGGHDS	344
Oy		374	EIDIGLSEGENLEKOMCOLAVIPPMUYDAOQAIFKLNNGMADPMKYKROQVANNWS	433
Db		345	E---LTTEE-KKHIROMS-----LMAEKN---DEEILD---	370
Oy		434	EOEKETEREFKEMQHPKN-FGLIASFLERTVAECVLYYTLTKNENYKSLVRSTRRKX	492
Db		371	FQSK-----PKSTPGLKNF-----GK	387
Oy		493	SQQQQQQQQ---QQQQQQQQQPMPRSSOREKOEKEKEKAEEKPEVENDKELLKE	548
Db		388	ATQSVMATSDVINKVNOQSOSLSGLTOEBLDIAATSAQESKOBELTODEIDHT-A	446
Oy		549	KTTDDSGEDNDEKEVASVKGRKRTANSOGRRKGRITRSMANESEAITP---QOSAEI	604
Db		447	RINMAAADPFQPAVNMMQJLSQAEKD-HIARIEAMAADDOSKIIVPPSIIOQDIEL	504
Oy		605	ASMEINESSR--WTEEMETAKKGCL-----EHGNMNAIAMWGSKTV---SCKNPF	652
Db		505	SMEEDMDHARIAMADEDFTHPVKGAVAPIYDENYSRDDALIDFPASTATVPVFAOPSFI	564
Oy		653	YNMYKKRONLEI-----LOQHKLKC---EKERNARRKKKAPAAASE	692
Db		565	ELISEEREHIKIAMAMEDNTPTVSTHPOQIEIELTEERKHIAH----IEAMASE	619
Oy		693	EAF-P-PUVEBEMEAS-GVSGN--BEEMYEAHALHASNEVPRCECGPATVNSSD	747
Db		620	DINAASPFIQCORAVATMPVPINHYHEBPITISOEBIDHIA-----RIITAMTEDN-TD	671
Oy		748	TESISPHTEAKODGONGPKRPATLGADGPDPGPPTRPTSRAPIETTPASEATGAPT	807
Db		672	LQTLPLPYOVK---QNEBP-----	686
Oy		808	PPPAPPSPAPPVYPKEKEKEETAAPVVEGEOKPAAEIALVDTGAAEYVSECT	867
Db		687	-----LSQEBIDHITRIAMANEDFGMYSIYSE-----HIPAVESELTP	725
Oy		868	EEABESPARGKAAEAAEATAEGALAAKKEGGSGAATKASGAPQODSDSATCSADEVD	927
Db		726	OEBLDHIAK-----ITGASNDISTLP-----	748
Oy		928	EAEGGDKNLRLSPRPILTPTGDPRANASPOKELD-LNKOKORAALIPPIOVTYVHEPP	985
Db		749	-----FTGKPSETSLTOEBLDHIAIRIAMASAEDYDV-XI-FEPPE	788
Oy		986	-----REBAATKAPAPAPP-----	1001
Db		789	LTOEBLDHIAKITAMAAOVOLPATORSITYHKVSLTPPPPEPSKHFBDULTQEBLDHIAK	848
Oy		1002	-----PONLQESDAPQOQSSPRGSRSPAPPADKEAFNA-----EAQ	1040
Db		849	IAMADMINDYTPTTAAPVQDEBERPTTESGSATSGADIPOEQBASGASGSPDNNAQC	908
Oy		1041	KLPDDEPCWTSGLPFFVPPPREVIKASPAAPDSATSYPAPGHLPLGLGHOTARPVLP-R	1098
Db		909	VL-----TSG-----FSPDRVTSAPALDITTEEPQGPIWA	937
Oy		1099	PPTITSNPPLISSAKPSV-----LERQICALSQG-----	1128
Db		938	QKVVSPTPSADSNAKRKSEYDIRKISLSEIROSESIDICKMYEQLSFTMRQSIHDEEDVG	997

Qy	1129	-----MSVOHV-----PYBNAKAPGPT-----MGLPLMDPKLAP	1163
Db	998	HEIRTDVEEPLFEYEBOLHFLBGIDVEOQHN-----PTTSAFCTGSIQEGEKXKG	1053
Qy	1164	FSYKQKOL-----SPRGAGPRESL--GVPTQOEA5V-LRGTAIGSVGCSITK-----G	1211
Db	1054	DDAVEOQKLENYEEETKSSSSREAPDDGETQRESLAQQPRLDLSGMYLKPFGG	1113
Qy	1212	IP-----TRVPSDAITTRGSIITHGPRADVLKGTITRIIGEDSPRLDR	1257
Db	1114	FLSNINDAIWKAKAGSQIOAAVPIPRSSSSNNIWNNVFSSSKSSTLSGTAPLK---	1170
Qy	1258	GREDSPRGHV-----IYEGKKGHVLSYEGMGSVYQCSKEDRSSSGPHETAAFKTY	1311
Db	1171	---SIBPQIGIPMDGISEEBERQIWSMAAADPDDSVANNVPRSTSSGNNIPAGMEDLS	1226
Qy	1312	DMMEGRVRAISASIEGLMGRAIPERRSHPHLKQHNIRGSIITQIPRSYEAQEDY-	1370
Db	1227	EAERKIMSWANAME--MGARPPR--SQIPTRSPVSSSIWELPRGLDLDDBDR	1283
Qy	1371	-----LRBAKLKREGRPPRRPSRLTEYKTOALGPLKLRHGLAATVK	1419
Db	1284	MKIMAWMAADMOVNRKIR--GRPPRPTSR-----IPGMGL--SBE	1326
Qy	1420	EAGRSIHEIPREELHNTPELRLAPRKESGITQGPLKYDTGASTGSKHDVRLIGS	1479
Db	1327	EQOKIMSWANAEIDSSSVITSTQPSKSPVAMQ-----QMP	1367
Qy	1480	PGRTPPRVNR-----LDVADARALERACEYESILKSRPTASSGSIARGA	1526
Db	1368	POQAPRIIPRGLEGLSDEBRTIMSWAEAE-----FEBSROVPRQPSRS-----	1414
Qy	1527	PVIVELKGRPSPLTYEDHAP--FAGHLPRGSFVTRKEPTRLQEGSLSSKASQDKL	1585
Db	1415	---PSFVNPQS-----FHPIPSPEPIVPRGLEDLSEEBOKIMSWANAEVBSKQL	1465
Qy	1586	TSTPREIAKSPHSTVPEHHPHPISPYEHLKGVGV-----DYKSHIPLAFDPTSIPRGI	1641
Db	1466	PS--RQSPSRSPSVAMIOAPAVPIIP-----SGMEDLPRAEQOKIMSWANAEIONFV	1515
Qy	1642	PIDAAAYULPHNLHPNTPYPLHYPRYLIRGYPALALENQITINDYITSO-----	1693
Db	1516	PSRSPNT---SMQPVYIPIH-----GLEDSEAE--RQKILSWANAEIDSAKIPS	1562
Qy	1694	-----QMHNTATMAQPADMLRGLSPRESLALNTAAGRGITIDSOV	1737
Db	1563	RSTSYSMPRLPQMSQPEITTGLEHSEADMEFG---RDBS--RSQVPIPRGLENLSEB	1617
Qy	1738	PHLPVLPPTPGIPATMDRLAULFLTA--POPFSSRHSSPLSPGQ-----	1781
Db	1618	EQOIN-----SWANAEISITPSGHIDPISLRPHGTGFKPAGVINEDELPERKQ	1671
Qy	1782	-----PYHLTKPTTSSSERE-----RD--RDREKRD-----RERE	1811
Db	1672	REESPLRREGVATSTYERELAMGBERMDGLIEDIIRIGAGSRDSDREVLHRRERD	1731
Qy	1812	KSLTL-----STTYENAPITWRPCTQSSSSSSSGGSSSRPASHAHQHSPISPRT	1867
Db	1732	PEHNTBEESTAVTVDVPSVSPTEVENPEKQTDPDFYTSDBRFALFIQOMEBEBAQLO	1791
Qy	1868	QDALOORPSVLHNTGAKGIIITAVEBSKPLVLAISTSSBPVAPATFPATHCPLGTLDG	1927
Db	1792	KQVYDEKPRMWE-----IVFDGDSSELPHQGFVNEBPTTK-----KTSDF	1831
Qy	1928	VYPTLMEPVL-LPKEAPVAPREPRADTGNAPLAKPRARSGLEBASPSKGSBPRPLVP	1966
Db	1832	DFPEKTEDEVKEPSELORIRVTKHNDVDMEIY-----DNVIATEA5SVQSR---Q	1881
Qy	1987	PVSGHA-----TARTPRAKULAVHNASDPAPRAPASDPNHKEQOSKXPSIOELELSL	2041
Db	1882	PVDSSETSUKRTIQRGPTK-----PPMIKITVEBETKS-----DDESSC	1922
Qy	2042	GYNHSSYSPGCVAVPSVPSLITNDKPLKHLLEILDKSHLEGELRPKQPGVYUKLGMEA	2104


```

QY 1011 APQPGSSPRKSSRPAPADK-----EAPAAEAQKLPFGDPDPCWTSGLPFVPPREVIXA 1065
Db 369 QPQPGPLVPPQQAQAPFGPGRGKGDGPGCTG--LPG--PSGRDGLPG-----415
QY 1066 SPNAPDPAFAVAPGHPLPLGLHDTAPVLPRLPTISNPPPL-----ISSAKH 1114
Db 416 -PGSD-----GPPGDP---GYTNGIYECQPPGDDGPPFGIIPQGPFGIEIGKQKG 464
QY 1115 PSVLEROI GAI-----SQGMSVQLHVPYSEHAKA-----PVGPTVGLPLPMDPKLAPFS 1165
Db 465 ESCLICIDYGRGPPGPPGCEIGFPQGPAGKAGDRGLPGRDGVAGVGPQCTPGLIQP 524
QY 1166 GVKOE-----QLSRGQAGPPESIGVPTAOEASVLRGTALGVSQGSITIGISTRPS 1219
Db 525 GAKGPPGFYDLRLKDKGDPFGPQPG-----MPRA--GS--PGRDGHPLPGPK--GS 575
QY 1220 DSATYRGSITHTGPAVLYKGTITRIIGEDSPSLDGRREDSLPKGHVYE-----GK 1274
Db 576 PGSVGLKE--RGPPGVGVFGPS-----KQDTGPPGPPGPPGAPIDDK 617
QY 1275 GHVLSYEGMSVYQCSKEDGRSSSGPPHETA PKTYDMMEGRVGRALISASIEGLMG-- 1332
Db 618 G-----QAGPPGPGSPGLPGKGP-----GKIYPLPGPPGAGGLPGSP 657
QY 1333 -----RAIPRKHSHHLKEOHNGISITQ--GIPRSVYEAQEDVLRREALKRE 1381
Db 658 GPPGQDGRGPPGTGRPLPGE-----KQAVGQPGI GPPGPPGKGVGLPD--MPP 710
QY 1382 GTPPPPP-----PSRLTAAYKTQ--ALGPLKLAHGLVATVEAGR--SHEIPREE- 1432
Db 711 GTPGRFENGRLPGANGVQKQKGPVGLPGLKGLPGLPGIPIPTPEKKSIGVPGVGHG 770
QY 1433 -----LRHTPELPLAPRLKEGSI--TGTPPLKYDTGASTGSK--KHDVSLIG 1478
Db 771 AIGPPGLGIRGEPPGPLP-----GSVGSPPGV--GIQPGARPGGQGPGLISG 820
QY 1479 SPG---RTPPVHLDVMA-----DARALERACTEESLKSPRTASSGGSIMAGAPVI 1529
Db 821 PPGIKGEGFPGLDMPGPKDKAQGLPGITQSGPLPGLPGQGAAPGIPGPPSGKE 880
QY 1530 VPGLKPPROS-----PLVEDHGAAPAGHLPGSPVPTMBE-----PTP 1567
Db 881 MGWMTPPQPGSPGVVGAAGLPGEGKDHGFPSSG--PKQDPBLKDKDVGLPKPGSMD 939
QY 1568 RLQEGSLSSSKASQDK-----LTSTPREIAK-----SPHSVYENHHP--- 1606
Db 940 KVDMSMKQKQDQEGKQIGPIGKGSRGDPGTGVPKQDQAGQPGQPGKDPGISG 999
QY 1607 -----PISPYHLARGVGVULYRSHITLADPTISIPGILPDA 1647
Db 1000 TPGABLPQPKGSVSGMLPGTPG--KGVPGIP-----GPGSP--GLPGDKGA 1045
QY 1648 AYYLPRHLAPNPTYPHLYPPLIRGYPTALLENQTIINDYITSQOMHNTATAMAORA 1707
Db 1046 KGEKQAGPPLGIGL-----RG-----EKG 1067
QY 1708 DM-LRGL--SPRESSLALNYAAGPRGIIDLQVPHLPVL-----VPPFGTATAMD 1756
Db 1068 DQGIAGPSPGPK-----GEGKSIGIPGMPPSPGLKSPGSGVGVPGSLGEGEKD 1119
QY 1757 RLAYLPTAPQ--PSSSRHSSPLSPGPHLTKPTTSSSRERDRDRERDREREKSL 1815
Db 1120 K--GLPGLDGLPGVKGAGLPETPG-----PTBPAGQKGPESDGLPGSAGEKEPGL 1170
QY 1816 TSTTVEHAPLWRPCTEOSGSSGSSG-----GGGSSSRPASHSHAHOSHISPRTOA 1870
Db 1171 PGRG-----PFPRAKDKGSKGAVGFPGLGAPPIIPSGKQEGMGPPGQGP 1222
QY 1871 LQQRPSVLHNT-----GMKGIITAVEBSKPTVLASTSTSSPVNPATPPATCPICG 1923
Db 1223 LPGSPG--HATGPKDGRGPG-----QPLGLPL-----PGWGPFG-----LPG 1260

```

```

QY 1924 TLQGVYPTIMEPVLLPKBAPVARPERPRADTGAFLAKPPARSGLEBSPSKGSEBRP 1983
Db 1261 -IDGVKGDKNP-----GMPAGVPBGKDPG--FGCMPI--GSGGITGSKDMP 1310
QY 1984 LVPPVSGHATTARTP-----AKNLAHHNSPDPAPAPASADPREKTQSKP 2030
Db 1311 GVPPGQPKGLPGLQGIKAGDQDQVPGAKL-----PGRPPFG-----P 1351
QY 2031 FSIQELRLSLGYHSSSPGEPVPSPSLTHDKGLPKHLELDKSLBEGELRK- 2089
Db 1352 YDIIRG--PGLPG--PBG-----PGLKGLQGLP-----GPKG 1381
QY 2090 QPVPYKGEAHLPHLRPLPESQSSPLLQTPAGVGHQVYLAQHISEVITQDYTR 2149
Db 1382 QQGVTLGVG-----LP--GPPGIPGDPGAPQKG-----1408
QY 2150 HHPOOLSAPLPAPLYSPGASCPLYDLRRPPSDLYLPPPDHAPARGSPHSEGGKRSPP 2209
Db 1409 ---EMGAPGTPGRGPGP-----PGPD-----GLPGSMGPPGTPSV 1442
QY 2210 NKTSLVGGEDGIEPVSPPEGMPERHSRAVYPLLRDG--EOTBPSMGSKSP--GNTS 2266
Db 1443 DHGFLVTHSQTIIDPPQPSGKTILYHG---YSLVYQGERAHQODLGTAGSCLRRFS 1498
QY 2267 QPAPFSLTBSNSAMVYSKKOEINKLNTNRNREPEYISQPTGEPNMPAITGTGLMT 2326
Db 1499 TMPPLFC-----NINNVNCFASRNDYSTWLSLTPREMPMSAPITGENTRP 1543
QY 2327 YRSQVQDHASTNNGLEAIRKALMKYQWEEBPPLSANAFNPLNASASLPAMPITAA 2386
Db 1544 FISRCVACEAPA-----MVAVHSTQIQLP--CPSGMSSIMIGSYFM-----1585
QY 2387 DGRSDHTLTPSGGGKAVSGRPSRKAASAPGLASD--RPPSVSVSHSGDCN 2440
Db 1586 -----HTSAGAGSGQALAS-----PSCLEFRSAPPIECHGRGTGN 1623

RESULT 97
725752
Hypothetical protein F45B4.4 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C.Accession: T25752
R.Wilson, R.
submitted to the EMBL Data Library, September 1996
A.Description: The sequence of C. elegans cosmid F45B4.
A.Reference number: 220082
A.Accession: T25752
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2361 <MUL>
A.Citobas-References: EMBL:U70852; PTDN:AA09135.1; GSPDB:GN00022; CESP:F45B4.4
A.Experimental source: strain Bristol N2; clone F45B4
A.Gene: CESP:F45B4.4
A.Map position: 4
A.Intons: 60/2; 111/2; 939/3; 977/2; 998/1; 1021/1; 1125/3; 1140/3; 1157/2; 1173/2; 131.

Query Match 2.4%; Score 319; DB 2; Length 2361;
Best Local Similarity 17.6%; Pred. No. 0.00068;
Matches 500; Conservative 355; Mismatches 1017; Indels 968; Gaps 118;

```

A: Molecule type: protein
 A: Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GIA>
 A: Note: The amino end of the mature form is blocked
 R: Soliman, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
 FEBS Lett. 225, 188-194, 1987
 A: Title: Complete primary structure of the alpha(1)-chain of human basement membrane (type IV) collagen
 A: Reference number: S00207; MUID:88083584; PMID:3691802
 A: Accession: S00207
 A: Molecule type: mRNA
 A: Residues: 244-530 <S013>
 A: Cross-references: EMBL:Y00706, NID:G29548, PIDD:CAA68698.1; PID:G29549
 R: Edle, J.A.; Goldik, R.; Mann, K.; Kuenn, K.
 EMBO J. 12, 4795-4802, 1993
 A: Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
 A: Reference number: S39614, MUID:94038963; PMID:8223488
 A: Accession: S39614
 A: Molecule type: protein
 A: Residues: 371-554 <EBL>
 R: Babel, W.; Glanville, R.W.
 Eur. J. Biochem. 143, 545-556, 1984
 A: Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid sequence
 A: Reference number: A02863; MUID:85003629; PMID:6434307
 A: Accession: A02863
 A: Molecule type: protein
 A: Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-1001
 A: Experimental source: placenta
 R: Glanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A: Title: Peptide fragments of human placental basement-membrane collagens showing interrupted
 A: Reference number: S16908; MUID:82005835; PMID:6792033
 A: Accession: S16908
 A: Molecule type: protein
 A: Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-1412
 R: MacKnight, R.S.; Benson, V.A.; Lovell, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A: Title: Isolation and characterization of pepsin-solubilized human basement membrane (type IV) collagen
 A: Reference number: S16910; MUID:84053346; PMID:6416291
 A: Accession: S16910
 A: Molecule type: protein
 A: Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-949
 A: Experimental source: placenta
 R: Philanthem, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; F.
 J. Biol. Chem. 260, 7681-7687, 1985
 A: Title: CDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
 A: Reference number: S01466; MUID:85207819; PMID:2581969
 A: Accession: S01466
 A: Molecule type: mRNA
 A: Residues: 1256-1669 <PIH>
 A: Cross-references: EMBL:M10940, NID:G180421, PIDD:AAAS2006.1; PID:G180424
 R: Brinker, J.M.; Gudus, L.U.; Loidl, H.R.; Wang, S.Y.; Rosenblum, U.; Ketallides, N.A.; Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
 A: Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
 A: Reference number: S16879; MUID:85216555; PMID:2582422
 A: Accession: S16879
 A: Molecule type: mRNA
 A: Residues: 1259-1663 <BRI>
 A: Cross-references: EMBL:M11315, NID:G180817, PIDD:AAAS2042.1; PID:G180818
 R: Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985
 A: Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1 type IV procollagen
 A: Reference number: A02864; MUID:85127033; PMID:2578961
 A: Accession: S19051
 A: Molecule type: protein
 A: Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X', 1555-1556, 'H', 1558-1559, 'H', 1561-1562, 'H', 1564-1565, 'H', 1567-1568, 'H', 1570-1571, 'H', 1573-1574, 'H', 1576-1577, 'H', 1579-1580, 'H', 1582-1583, 'H', 1585-1586, 'H', 1588-1589, 'H', 1591-1592, 'H', 1594-1595, 'H', 1597-1598, 'H', 1600-1601, 'H', 1603-1604, 'H', 1606-1607, 'H', 1609-1610, 'H', 1612-1613, 'H', 1615-1616, 'H', 1618-1619, 'H', 1621-1622, 'H', 1624-1625, 'H', 1627-1628, 'H', 1630-1631, 'H', 1633-1634, 'H', 1636-1637, 'H', 1639-1640, 'H', 1642-1643, 'H', 1645-1646, 'H', 1648-1649, 'H', 1651-1652, 'H', 1654-1655, 'H', 1657-1658, 'H', 1660-1661, 'H', 1663-1664, 'H', 1666-1667, 'H', 1669-1670, 'H', 1672-1673, 'H', 1675-1676, 'H', 1678-1679, 'H', 1681-1682, 'H', 1684-1685, 'H', 1687-1688, 'H', 1690-1691, 'H', 1693-1694, 'H', 1696-1697, 'H', 1699-1700, 'H', 1702-1703, 'H', 1705-1706, 'H', 1708-1709, 'H', 1711-1712, 'H', 1714-1715, 'H', 1717-1718, 'H', 1720-1721, 'H', 1723-1724, 'H', 1726-1727, 'H', 1729-1730, 'H', 1732-1733, 'H', 1735-1736, 'H', 1738-1739, 'H', 1741-1742, 'H', 1744-1745, 'H', 1747-1748, 'H', 1750-1751, 'H', 1753-1754, 'H', 1756-1757, 'H', 1759-1760, 'H', 1762-1763, 'H', 1765-1766, 'H', 1768-1769, 'H', 1771-1772, 'H', 1774-1775, 'H', 1777-1778, 'H', 1780-1781, 'H', 1783-1784, 'H', 1786-1787, 'H', 1789-1790, 'H', 1792-1793, 'H', 1795-1796, 'H', 1798-1799, 'H', 1801-1802, 'H', 1804-1805, 'H', 1807-1808, 'H', 1810-1811, 'H', 1813-1814, 'H', 1816-1817, 'H', 1819-1820, 'H', 1822-1823, 'H', 1825-1826, 'H', 1828-1829, 'H', 1831-1832, 'H', 1834-1835, 'H', 1837-1838, 'H', 1840-1841, 'H', 1843-1844, 'H', 1846-1847, 'H', 1849-1850, 'H', 1852-1853, 'H', 1855-1856, 'H', 1858-1859, 'H', 1861-1862, 'H', 1864-1865, 'H', 1867-1868, 'H', 1870-1871, 'H', 1873-1874, 'H', 1876-1877, 'H', 1879-1880, 'H', 1882-1883, 'H', 1885-1886, 'H', 1888-1889, 'H', 1891-1892, 'H', 1894-1895, 'H', 1897-1898, 'H', 1900-1901, 'H', 1903-1904, 'H', 1906-1907, 'H', 1909-1910, 'H', 1912-1913, 'H', 1915-1916, 'H', 1918-1919, 'H', 1921-1922, 'H', 1924-1925, 'H', 1927-1928, 'H', 1930-1931, 'H', 1933-1934, 'H', 1936-1937, 'H', 1939-1940, 'H', 1942-1943, 'H', 1945-1946, 'H', 1948-1949, 'H', 1951-1952, 'H', 1954-1955, 'H', 1957-1958, 'H', 1960-1961, 'H', 1963-1964, 'H', 1966-1967, 'H', 1969-1970, 'H', 1972-1973, 'H', 1975-1976, 'H', 1978-1979, 'H', 1981-1982, 'H', 1984-1985, 'H', 1987-1988, 'H', 1990-1991, 'H', 1993-1994, 'H', 1996-1997, 'H', 1999-2000, 'H', 2002-2003, 'H', 2005-2006, 'H', 2008-2009, 'H', 2011-2012, 'H', 2014-2015, 'H', 2017-2018, 'H', 2020-2021, 'H', 2023-2024, 'H', 2026-2027, 'H', 2029-2030, 'H', 2032-2033, 'H', 2035-2036, 'H', 2038-2039, 'H', 2041-2042, 'H', 2044-2045, 'H', 2047-2048, 'H', 2050-2051, 'H', 2053-2054, 'H', 2056-2057, 'H', 2059-2060, 'H', 2062-2063, 'H', 2065-2066, 'H', 2068-2069, 'H', 2071-2072, 'H', 2074-2075, 'H', 2077-2078, 'H', 2080-2081, 'H', 2083-2084, 'H', 2086-2087, 'H', 2089-2090, 'H', 2092-2093, 'H', 2095-2096, 'H', 2098-2099, 'H', 2101-2102, 'H', 2104-2105, 'H', 2107-2108, 'H', 2110-2111, 'H', 2113-2114, 'H', 2116-2117, 'H', 2119-2120, 'H', 2122-2123, 'H', 2125-2126, 'H', 2128-2129, 'H', 2131-2132, 'H', 2134-2135, 'H', 2137-2138, 'H', 2140-2141, 'H', 2143-2144, 'H', 2146-2147, 'H', 2149-2150, 'H', 2152-2153, 'H', 2155-2156, 'H', 2158-2159, 'H', 2161-2162, 'H', 2164-2165, 'H', 2167-2168, 'H', 2170-2171, 'H', 2173-2174, 'H', 2176-2177, 'H', 2179-2180, 'H', 2182-2183, 'H', 2185-2186, 'H', 2188-2189, 'H', 2191-2192, 'H', 2194-2195, 'H', 2197-2198, 'H', 2200-2201, 'H', 2203-2204, 'H', 2206-2207, 'H', 2209-2210, 'H', 2212-2213, 'H', 2215-2216, 'H', 2218-2219, 'H', 2221-2222, 'H', 2224-2225, 'H', 2227-2228, 'H', 2230-2231, 'H', 2233-2234, 'H', 2236-2237, 'H', 2239-2240, 'H', 2242-2243, 'H', 2245-2246, 'H', 2248-2249, 'H', 2251-2252, 'H', 2254-2255, 'H', 2257-2258, 'H', 2260-2261, 'H', 2263-2264, 'H', 2266-2267, 'H', 2269-2270, 'H', 2272-2273, 'H', 2275-2276, 'H', 2278-2279, 'H', 2281-2282, 'H', 2284-2285, 'H', 2287-2288, 'H', 2290-2291, 'H', 2293-2294, 'H', 2296-2297, 'H', 2299-2300, 'H', 2302-2303, 'H', 2305-2306, 'H', 2308-2309, 'H', 2311-2312, 'H', 2314-2315, 'H', 2317-2318, 'H', 2320-2321, 'H', 2323-2324, 'H', 2326-2327, 'H', 2329-2330, 'H', 2332-2333, 'H', 2335-2336, 'H', 2338-2339, 'H', 2341-2342, 'H', 2344-2345, 'H', 2347-2348, 'H', 2350-2351, 'H', 2353-2354, 'H', 2356-2357, 'H', 2359-2360, 'H', 2362-2363, 'H', 2365-2366, 'H', 2368-2369, 'H', 2371-2372, 'H', 2374-2375, 'H', 2377-2378, 'H', 2380-2381, 'H', 2383-2384, 'H', 2386-2387, 'H', 2389-2390, 'H', 2392-2393, 'H', 2395-2396, 'H', 2398-2399, 'H', 2401-2402, 'H', 2404-2405, 'H', 2407-2408, 'H', 2410-2411, 'H', 2413-2414, 'H', 2416-2417, 'H', 2419-2420, 'H', 2422-2423, 'H', 2425-2426, 'H', 2428-2429, 'H', 2431-2432, 'H', 2434-2435, 'H', 2437-2438, 'H', 2440-2441, 'H', 2443-2444, 'H', 2446-2447, 'H', 2449-2450, 'H', 2452-2453, 'H', 2455-2456, 'H', 2458-2459, 'H', 2461-2462, 'H', 2464-2465, 'H', 2467-2468, 'H', 2470-2471, 'H', 2473-2474, 'H', 2476-2477, 'H', 2479-2480, 'H', 2482-2483, 'H', 2485-2486, 'H', 2488-2489, 'H', 2491-2492, 'H', 2494-2495, 'H', 2497-2498, 'H', 2500-2501, 'H', 2503-2504, 'H', 2506-2507, 'H', 2509-2510, 'H', 2512-2513, 'H', 2515-2516, 'H', 2518-2519, 'H', 2521-2522, 'H', 2524-2525, 'H', 2527-2528, 'H', 2530-2531, 'H', 2533-2534, 'H', 2536-2537, 'H', 2539-2540, 'H', 2542-2543, 'H', 2545-2546, 'H', 2548-2549, 'H', 2551-2552, 'H', 2554-2555, 'H', 2557-2558, 'H', 2560-2561, 'H', 2563-2564, 'H', 2566-2567, 'H', 2569-2570, 'H', 2572-2573, 'H', 2575-2576, 'H', 2578-2579, 'H', 2581-2582, 'H', 2584-2585, 'H', 2587-2588, 'H', 2590-2591, 'H', 2593-2594, 'H', 2596-2597, 'H', 2599-2600, 'H', 2602-2603, 'H', 2605-2606, 'H', 2608-2609, 'H', 2611-2612, 'H', 2614-2615, 'H', 2617-2618, 'H', 2620-2621, 'H', 2623-2624, 'H', 2626-2627, 'H', 2629-2630, 'H', 2632-2633, 'H', 2635-2636, 'H', 2638-2639, 'H', 2641-2642, 'H', 2644-2645, 'H', 2647-2648, 'H', 2650-2651, 'H', 2653-2654, 'H', 2656-2657, 'H', 2659-2660, 'H', 2662-2663, 'H', 2665-2666, 'H', 2668-2669, 'H', 2671-2672, 'H', 2674-2675, 'H', 2677-2678, 'H', 2680-2681, 'H', 2683-2684, 'H', 2686-2687, 'H', 2689-2690, 'H', 2692-2693, 'H', 2695-2696, 'H', 2698-2699, 'H', 2701-2702, 'H', 2704-2705, 'H', 2707-2708, 'H', 2710-2711, 'H', 2713-2714, 'H', 2716-2717, 'H', 2719-2720, 'H', 2722-2723, 'H', 2725-2726, 'H', 2728-2729, 'H', 2731-2732, 'H', 2734-2735, 'H', 2737-2738, 'H', 2740-2741, 'H', 2743-2744, 'H', 2746-2747, 'H', 2749-2750, 'H', 2752-2753, 'H', 2755-2756, 'H', 2758-2759, 'H', 2761-2762, 'H', 2764-2765, 'H', 2767-2768, 'H', 2770-2771, 'H', 2773-2774, 'H', 2776-2777, 'H', 2779-2780, 'H', 2782-2783, 'H', 2785-2786, 'H', 2788-2789, 'H', 2791-2792, 'H', 2794-2795, 'H', 2797-2798, 'H', 2800-2801, 'H', 2803-2804, 'H', 2806-2807, 'H', 2809-2810, 'H', 2812-2813, 'H', 2815-2816, 'H', 2818-2819, 'H', 2821-2822, 'H', 2824-2825, 'H', 2827-2828, 'H', 2830-2831, 'H', 2833-2834, 'H', 2836-2837, 'H', 2839-2840, 'H', 2842-2843, 'H', 2845-2846, 'H', 2848-2849, 'H', 2851-2852, 'H', 2854-2855, 'H', 2857-2858, 'H', 2860-2861, 'H', 2863-2864, 'H', 2866-2867, 'H', 2869-2870, 'H', 2872-2873, 'H', 2875-2876, 'H', 2878-2879, 'H', 2881-2882, 'H', 2884-2885, 'H', 2887-2888, 'H', 2890-2891, 'H', 2893-2894, 'H', 2896-2897, 'H', 2899-2900, 'H', 2902-2903, 'H', 2905-2906, 'H', 2908-2909, 'H', 2911-2912, 'H', 2914-2915, 'H', 2917-2918, 'H', 2920-2921, 'H', 2923-2924, 'H', 2926-2927, 'H', 2929-2930, 'H', 2932-2933, 'H', 2935-2936, 'H', 2938-2939, 'H', 2941-2942, 'H', 2944-2945, 'H', 2947-2948, 'H', 2950-2951, 'H', 2953-2954, 'H', 2956-2957, 'H', 2959-2960, 'H', 2962-2963, 'H', 2965-2966, 'H', 2968-2969, 'H', 2971-2972, 'H', 2974-2975, 'H', 2977-2978, 'H', 2980-2981, 'H', 2983-2984, 'H', 2986-2987, 'H', 2989-2990, 'H', 2992-2993, 'H', 2995-2996, 'H', 2998-2999, 'H', 3001-3002, 'H', 3004-3005, 'H', 3007-3008, 'H', 3010-3011, 'H', 3013-3014, 'H', 3016-3017, 'H', 3019-3020, 'H', 3022-3023, 'H', 3025-3026, 'H', 3028-3029, 'H', 3031-3032, 'H', 3034-3035, 'H', 3037-3038, 'H', 3040-3041, 'H', 3043-3044, 'H', 3046-3047, 'H', 3049-3050, 'H', 3052-3053, 'H', 3055-3056, 'H', 3058-3059, 'H', 3061-3062, 'H', 3064-3065, 'H', 3067-3068, 'H', 3070-3071, 'H', 3073-3074, 'H', 3076-3077, 'H', 3079-3080, 'H', 3082-3083, 'H', 3085-3086, 'H', 3088-3089, 'H', 3091-3092, 'H', 3094-3095, 'H', 3097-3098, 'H', 3100-3101, 'H', 3103-3104, 'H', 3106-3107, 'H', 3109-3110, 'H', 3112-3113, 'H', 3115-3116, 'H', 3118-3119, 'H', 3121-3122, 'H', 3124-3125, 'H', 3127-3128, 'H', 3130-3131, 'H', 3133-3134, 'H', 3136-3137, 'H', 3139-3140, 'H', 3142-3143, 'H', 3145-3146, 'H', 3148-3149, 'H', 3151-3152, 'H', 3154-3155, 'H', 3157-3158, 'H', 3160-3161, 'H', 3163-3164, 'H', 3166-3167, 'H', 3169-3170, 'H', 3172-3173, 'H', 3175-3176, 'H', 3178-3179, 'H', 3181-3182, 'H', 3184-3185, 'H', 3187-3188, 'H', 3190-3191, 'H', 3193-3194, 'H', 3196-3197, 'H', 3199-3200, 'H', 3202-3203, 'H', 3205-3206, 'H', 3208-3209, 'H', 3211-3212, 'H', 3214-3215, 'H', 3217-3218, 'H', 3220-3221, 'H', 3223-3224, 'H', 3226-3227, 'H', 3229-3230, 'H', 3232-3233, 'H', 3235-3236, 'H', 3238-3239, 'H', 3241-3242, 'H', 3244-3245, 'H', 3247-3248, 'H', 3250-3251, 'H', 3253-3254, 'H', 3256-3257, 'H', 3259-3260, 'H', 3262-3263, 'H', 3265-3266, 'H', 3268-3269, 'H', 3271-3272, 'H', 3274-3275, 'H', 3277-3278, 'H', 3280-3281, 'H', 3283-3284, 'H', 3286-3287, 'H', 3289-3290, 'H', 3292-3293, 'H', 3295-3296, 'H', 3298-3299, 'H', 3301-3302, 'H', 3304-3305, 'H', 3307-3308, 'H', 3310-3311, 'H', 3313-3314, 'H', 3316-3317, 'H', 3319-3320, 'H', 3322-3323, 'H', 3325-3326, 'H', 3328-3329, 'H', 3331-3332, 'H', 3334-3335, 'H', 3337-3338, 'H', 3340-3341, 'H', 3343-3344, 'H', 3346-3347, 'H', 3349-3350, 'H', 3352-3353, 'H', 3355-3356, 'H', 3358-3359, 'H', 3361-3362, 'H', 3364-3365, 'H', 3367-3368, 'H', 3370-3371, 'H', 3373-3374, 'H', 3376-3377, 'H', 3379-3380, 'H', 3382-3383, 'H', 3385-3386, 'H', 3388-3389, 'H', 3391-3392, 'H', 3394-3395, 'H', 3397-3398, 'H', 3400-3401, 'H', 3403-3404, 'H', 3406-3407, 'H', 3409-3410, 'H', 3412-3413, 'H', 3415-3416, 'H', 3418-3419, 'H', 3421-3422, 'H', 3424-3425, 'H', 3427-3428, 'H', 3430-3431, 'H', 3433-3434, 'H', 3436-3437, 'H', 3439-3440, 'H', 3442-3443, 'H', 3445-3446, 'H', 3448-3449, 'H', 3451-3452, 'H', 3454-3455, 'H', 3457-3458, 'H', 3460-3461, 'H', 3463-3464, 'H', 3466-3467, 'H', 3469-3470, 'H', 3472-3473, 'H', 3475-3476, 'H', 3478-3479, 'H', 3481-3482, 'H', 3484-3485, 'H', 3487-3488, 'H', 3490-3491, 'H', 3493-3494, 'H', 3496-3497, 'H', 3499-3500, 'H', 3502-3503, 'H', 3505-3506, 'H', 3508-3509, 'H', 3511-3512, 'H', 3514-3515, 'H', 3517-3518, 'H', 3520-3521, 'H', 3523-3524, 'H', 3526-3527, 'H', 3529-3530, 'H', 3532-3533, 'H', 3535-3536, 'H', 3538-3539, 'H', 3541-3542, 'H', 3544-3545, 'H', 3547-3548, 'H', 3550-3551, 'H', 3553-3554, 'H', 3556-3557, 'H', 3559-3560, 'H', 3562-3563, 'H', 3565-3566, 'H', 3568-3569, 'H', 3571-3572, 'H', 3574-3575, 'H', 3577-3578, 'H', 3580-3581, 'H', 3583-3584, 'H', 3586-3587, 'H', 3589-3590, 'H', 3592-3593, 'H', 3595-3596, 'H', 3598-3599, 'H', 3601-3602, 'H', 3604-3605, 'H', 3607-3608, 'H', 3610-3611, 'H', 3613-3614, 'H', 3616-3617, 'H', 3619-3620, 'H', 3622-3623, 'H', 3625-3626, 'H', 3628-3629, 'H', 3631-3632, 'H', 3634-3635, 'H', 3637-3638, 'H', 3640-3641, 'H', 3643-3644, 'H', 3646-3647, 'H', 3649-3650, 'H', 3652-3653, 'H', 3655-3656, 'H', 3658-3659, 'H', 3661-3662, 'H', 3664-3665, 'H', 3667-3668, 'H', 3670-3671, 'H', 3673-3674, 'H', 3676-3677, 'H', 3679-3680, 'H', 3682-3683, 'H', 3685-3686, 'H', 3688-3689, 'H', 3691-3692, 'H', 3694-3695, 'H', 3697-3698, 'H', 3700-3701, 'H', 3703-3704, 'H', 3706-3707, 'H', 3709-3710, 'H', 3712-3713, 'H', 3715-3716, 'H', 3718-3719, 'H', 3721-3722, 'H', 3724-3725, 'H', 3727-3728, 'H', 3730-3731, 'H', 3733-3734, 'H', 3736-3737, 'H', 3739-3740, 'H', 3742-3743, 'H', 3745-3746, 'H', 3748-3749, 'H', 3751-3752, 'H', 3754-3755, 'H', 3757-3758, 'H', 3760

```

Db      726 EQAVEADAGPTGANE-----DDSDVPAVVPVLESDAVNERCEMEQQAOKSMEQ 773
Qy      843 QKPPAEELAVDTGKAE-----EVPKSEC-- 866
Db      774 PEQKAATEVSKELSESQVHMMAAAAVADGTAAITIEERSPSMISASVTEPLEQVBAAL 833
Qy      867 -TEE-----AEGCP-----AKGDAEAAEATAEALAEKKEGGSGAATTAKSQA 911
Db      834 LEELELEVEVIAEEBPPVTEPLPENREARQDVVSEALTP-----AVTAATFAG 885
Qy      912 PQDSDSATSCADEVDAEGDKNLLSPPSLTPPTGDPANASPOKPL-----DLKOL 966
Db      886 P-----LGAEBETASAAEBETTEMVANSQITDSQDTTEATPVEVEGQVDIEQ 937
Qy      967 KQRAAIPPIQVTKVNEPPREDAFTKPAAPPAPPPONLOPESDAPQOGSSPRKSRSP 1026
Db      938 ERRTQEVLAQVAEKVESQ-----LPGTGGEVDILO-----PVQRAEAERPEQAE 984
Qy      1027 APPADKE---APAAEQKLPBDPPCWTSGLPFPVPPREVITKASPAAPPSAFSTAPPGHP 1083
Db      985 ASGKKEEDVVLKVDQAQAKTEP--FTQG-----KVGQT-----TPESFEKAP--- 1026
Qy      1084 LPLGLHDARVLP RPPTISNPPLISSAKHPSVLERQIGALISQMSYQLHVPYSEHAKA 1143
Db      1027 -----QVTESESSE----- 1036
Qy      1144 PVGPVTMGLPLPMDPKCLAPFSGVKQEQLSPRGQAPPELSGVPAGQASVLRGTALGSV 1203
Db      1037 -----LVTTQCEFTLA-----GVKSGQVWME-QALIPDSVETPDSFTDSTPAVDPA 1084
Qy      1204 PGGSTTKIPESTRVPSDAITRRGSIHTGTPADVLYKGTITRIIGEDSPRLDRREBSL 1263
Db      1085 PGTQKQDEIVIHENEVA---SGTQSGTEAEV-----PAOKER---DPA 1125
Qy      1264 PKGHVTEYGEKK-----GHVLSYEGMSVYQCSKEDGSSGPPHETAPKRTYDMM 1314
Db      1126 PSSFPQFQETKEQSKMETLEHTDKVESEVETYSILSKTEGTOEA---DQVADETKXV- 1180
Qy      1315 EGRVGRALISASIEGLMGRAIPPERHSPHLLKEQHHRIGSTIGIPRSYVEAOEDYLRE 1374
Db      1181 -----PFEGL-----EGSIDTGTIVREKTEVALKGE 1209
Qy      1375 -----AKLLKREGTPPPPPSRDL---TEAYTQALGPKLKAHNGLVATYKE 1420
Db      1210 GTEAECKDDALELQSHAKSPSPVEREMVQVREKTEA-----EPTH----- 1254
Qy      1421 AGRSIHEIPRELRTPPEPLAPRPLKESGITQGTPLKVDGASTGSKGDVRSLLGSP 1480
Db      1255 -----VNEKLEHETAVTVSEVSKOLLQTVNVP-----IDGAK--EVSSLGSP 1298
Qy      1481 GRTEPPVAPLVDMDARALERAQYSELSKSRPG---TASSGSGIARGAPVYELGKP 1536
Db      1289 -----PFCIGQF-----EAVCTKIQVSSSEASFTLTAABEKKVL--GETANILETGS- 1344
Qy      1537 ROSPLITYEDHGAFFAGHLPKRGSPVTKRETPRLQSGSLSSSAQSDRLKLTSPRIASAP 1596
Db      1345 -----TLR-----PAGAHVLEB-----KSSBKND--FAAPPGEDA--- 1374
Qy      1597 HSTVEHPHPISPYEHLRGVSGVDLYRSHIPLAFDPTISRGIPRLDAAAYVLPRIILA 1656
Db      1375 ---VF-----TGPDCAKSTPVIVSATV--KKGILSSD----- 1401
Qy      1657 PNFTYPLHYPLRYNGYDPTALENRQTIINDYITSQOMHNTATAMAGRADMLGLSPR 1716
Db      1402 -----LEG-EKTSLKWKMSDEVDQVACQEVKVALEIDLEPENGIILETKX 1447
Qy      1717 ESSSLANPAAGRGIIIDLSQVHLEVLVLPPTGTAITMDRLAYL--PRAPOFSSRRHSS 1775
Db      1448 SSKLVQN-----IIO-----TAVDQVTEETEAATEMLTSELQO 1481
Qy      1776 PLSPGPGHLYTKPTTSSSERDRDRERDREREKSLTSTTVEHAPIRRPTEQSS 1835

```

```

Db      1482 -----AHVTKADS-----QDAGQETKEKEEPEPLASQO---DETPITSAKESES 1522
Qy      1836 GSSGSGGGGSSSRPASHAHQHPISPTQALQORPSVLNMTGKGIITAVEPEKP 1895
Db      1523 TAVQA-----HSDISKMSASE-----KMTVEVE----- 1549
Qy      1896 TLRKSTSSPVRAATFPATHCPLGTLGVPVPLTMEPVLTPKE-----APVAPRP 1950
Db      1550 -----GST-----VNDQLEEVVLPSEEGGAGTKSVPE- 1579
Qy      1951 PRADTGHAFLLAKPPARGSLBPASSPSKGSBEPPLVPPVSGHATIAFTPAKNLAPHHASPD 2010
Db      1580 ---DDGHALLAERLEKSLVED--KEDEKGD---VDDPENQVSLADTDASG--GLTKESP 1631
Qy      2011 PPAPPASADPHREKTKQSPSTIOELRLSLGTHGSSISPRGCVPEVSVSSPLTHDKL 2070
Db      1632 TNGP-----KQKKEQDA---QEVLEQEGKHVSES-----DKAI 1661
Qy      2071 -PKHLEELDKSHLE 2083
Db      1662 TPQAQELQKQERE 1675

```

RESULT 96

```

CGHUB
collagen alpha 1(IV) chain precursor - human
NAlternate names: procollagen alpha 1(IV) chain
CSpecies: Homo sapiens (man)
CDate: 28-May-1986 #sequence revision 31-Dec-1992 #text change 07-Dec-1999
CAccession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
J.Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A>Title: Structural organization of the gene for the alpha-1 chain of human type IV coll.
A>Reference number: S16876; MUID:89340433; PMID:2701944
A>Accession: S16876
A>Status: nucleic acid sequence not shown; translation not shown
A>Molecule type: DNA
A>Residues: 1-1669 <S011>
A>Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AAA53098.1; PID:G180803
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988
J.Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A>Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A>Reference number: A92680; MUID:89034231; PMID:3182844
A>Accession: A32117
A>Molecule type: DNA
A>Residues: 1-28 <S012>
A>Cross-references: EMBL:J04217; NID:G180759; PIDN:AAA53097.1; PID:G553233
R.Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A>Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A>Reference number: S02738; MUID:89030632; PMID:2846280
A>Accession: S02738
A>Status: translation not shown
A>Molecule type: DNA
A>Residues: 1-6,'L',8-28 <POB>
A>Cross-references: EMBL:X12784; NID:G30072
R.Birzel, D.; Oberhauser, I.; Dieinger, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A>Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A>Reference number: S00048; MUID:88029471; PMID:3311751
A>Accession: S00048
A>Molecule type: mRNA
A>Residues: 1-318,'A',320-944 <BRA1>
A>Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067
A>Accession: S25826
A>Molecule type: protein
A>Residues: 271-318,'A',320-554 <BRA2>
R.Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risceli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A>Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (75
A>Reference number: A23115; MUID:86004708; PMID:4043082
A>Accession: A23115

```


Db 3167 DLDHTPIA-----DHANVYHGDSSRIEKHASP-----VA 3196
Qy 1678 ALENROTIINDYITSQMHNTATAMAQRADMLRGLSPRESSLALNVAAPRG-IIDLSQ 1736
Db 3197 TLEKKQKKVK-----SPKEKVDVPAPEKKEIRLARVTEPEAEIIVDPV 3243
Qy 1737 VPH-----LPVL-----VPTTGTATMDRLAIVLPTAQPF-----SSRH 1772
Db 3244 VKETKEKKDKSKLPFTGFPFSKTKTKTKTGYPTEAAYTGEIDTDRTHDLEGTSEH 3303
Qy 1773 SSSPLSPGPHLTPTTSSSR-----ERDRD-----R 1802
Db 3304 GEHPAFASSTITTEADPDHEKVIKPSRPHMSRMRHGEDVVEKPEGYILSSRTY 3363
Qy 1803 ERDREREKSLTSTTVEHAPIRPTEQSSGSSGSSGSSSR-----1850
Db 3364 EGPDNMKPIGELLETPLDHSQVYH-----NGSSGKLEKRSKEKIKTPSPERVPS 3417
Qy 1851 -----PASHSH-----AHQHSISPTQ-----DALQR-----PSVLAHTGM 1883
Db 3418 EIKLVARIQPLEHVEDSEISKLSFVKDRSRAPSPSMFHROSQKQGYEISPLYEGNLDV 3477
Qy 1884 KGITAVEPSKPTVLSTSTSSPVYPAATPPATPCP-----1920
Db 3478 TGRAAEVA-NPIEVKSVPEYHFKQEVSAQHPBESSQPRFRLIARMRHGEDDEIEKPD 3536
Qy 1921 -----LGGLDGVPTLMEPVLLPKREAPRVARPERPRAD-----TGHAFIA---K 1962
Db 3537 AYKFVTEVEGPDLDIRPV-----VELAQPISDHSQVYHNGSSWKKEKNIK 3583
Qy 1963 PPAASGLEPASPSPKSGSEPRPLVPVYSGHATITARPAAKLAPHNASDP-----PAP---2015
Db 3584 PP-----KPVNTKKGSLPEPEISEVHEIRLITRVASSSEBEEDVJIPASTVPKPPRL 3638
Qy 2016 -----ASADPHREKTQSKP-----FSIOELELSIG-HGS-SYSP-----2050
Db 3639 RFLSIGKKSPTSPESEFPECBOYDQPDVLTIRSSELEVMPLYHOSVPEYSPKFTPVKPV 3698
Qy 2051 -----EGVEPV-----SPVSPSPLT-----HDKG 2069
Db 3699 DKADERKKTIVYFRHEGEDFADKPEAVGFSSDVTGSLNEISRSELEHAIIHDS 3758
Qy 2070 LPKH-----LELDKSHLEGELPKQ-----2090
Db 3759 QVTHDGLSYKIDNKKOKSKDRTELEBPTISRLVAVLPQOEASDGSBPKDKDKGFRFR 3818
Qy 2091 -----PGPVKLGGEAHLPHL-----RPLPESQPESSPLIOTAPGVK 2127
Db 3819 FFSKASPKSGPYEPTPEGPIDVTGRVESLEHVPLOSREFPTTSPKKS--QEKPSBK 3875
Qy 2128 GHORVVTIAOH--ISEVITODYTRHHPOLSAPLPAPLYSPGAS-----CEVLDR 2177
Db 3876 SSERVEVERKRYHLARI-----RHGDEBETVFN-DTYGFASTYDGLBETSKSVYE 3928
Qy 2178 RPPSLULYPPPDHGAAPAGSPHSEGGKSPPEPKTSVLGGGEGDGLIEPVSPPR--GMTBPG 2235
Db 3929 ETPISVYNNVHHGSGISIVEKQKRAKISABEK-----DDDENVPKPSRFSPLRG 3981
Qy 2236 HSRSAVYVLLYRDGQTEPSRWGSKSPGNTSQPAPAFSKLTES-----NSAMWKSKEOIN 2291
Db 3982 OKRVLSYSTLYSG-----PLETTRDLLENIPISIVYHSSDOHAKRYLI 4029
Qy 2292 KKLANTHNREPEYNIQPGTEIFNMPAITGTGLM--TYRQAQVOE-----HASTN 2339
Db 4030 ARIR-HAGDEDEVAHKLHPVYSFPAYQSGILEKTLCKEVDVPIGRFARVYHSGSS 4088
Qy 2340 MGLEIIRKALMGKIDQMBESPLSANAFLNABASLPAAMPITAAOGRSD-----2391
Db 4089 GGLIWNLLKNRNSRASKMDEPKRRKQKQDKIKLILAVLPMKTEPKSDRFEKATKE 4148
Qy 2392 -----HTLTSGGGKAKVSGRPSRRKAKSPAPGLAGSDRPPSVSVHSEGD 2439

Db 4149 EPLREAEVLNSTQSEAQNISTVSIPLSLTYDVP-----RPARIEGTYGGDC 4197
RESULT 94
A45344
Immediate-early protein - suid herpesvirus 1 (strain Kaplan)
C:Species: suid herpesvirus 1
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: A45344
R:Vilek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented op-
A:Reference number: A45344; MUID:91021039; PMID:2171211
A:Accession: A45344
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-1446 <VLC>
A:Cross-references: GB:M4651; NID:G334070; PIDN:AAA47470.1; PID:G334071
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation
Query Match 2.4%; Score 323; DB 1; Length 1446;
Best Local Similarity 20.5%; Pred. No. 0.00029;
Matches 344; Conservative 158; Mismatches 590; Indels 582; Gaps 78;
Qy 686 APPAASEAAAPPVVEDEMEASVSGNEEMVEEALHNASGNEVPRGEGSGPATVNS 745
Db 72 AAAAGATPPPPPAQOQOQPRRG-SGEIVLVDDEDEDEDEPGS--PAAGSPGRALHGS 128
Qy 746 SDTESIPSPHTEAADTQONGPKPP-----ATLGADGP-----PGP 782
Db 129 EHGHLVIGPRSRAG-----SGPRPTPALAAALAGAGCGRSPSAASPASSGSGP 183
Qy 783 PTPPRRTSRA--PI-EPTPASE-ATGAPTPPA-----PPSPAPPVVP- 823
Db 184 SAAPRWGPAGDPVGPBGPAPRPTPAPPAQAAVAAPARGPASPAAGVSAFG 243
Qy 824 -----KEKEKETAA-----PVYEGEGQKPPAAEELAVD 854
Db 244 GCGAPSAAGDRHHHOREPLDPPAARLDRPPLGARSVSNPNSNSNTTVAVE 303
Qy 855 T-----GKAEE-----PVKSECTEAEERG-----PAKGDAEAAEATA 887
Db 304 TVAKCPEDDEGLAGGGAAPQORPPRRRAGGALRGKGFSSSSGSDSLSPARS 363
Qy 888 EGALKAERKEGSGRATTAKSGAPQDSATCSADEVDABEGDKNRLISPRSLTP 947
Db 364 PSAPRA--PAAARRASASSSSSSSSSSSSSSSSSSSSSGEERDEG-----VRP-----407
Qy 948 TGDPRAANSPOKPLDKOLKQRAAIIPIQVTKYHPRREDAAPTKAPRAPPPQNLQ 1007
Db 408 GAPIARAGP-----PSPPAPAAAPRASASA 434
Qy 1008 ESDAPQCGSSPRGKRSRPA--PPADKEAPFAEAQKLPDPCWTSGLFPVPREVIRAS 1066
Db 435 SSSAASAPAP--EPAPPPRRKRSTNNHLSIMADGPPIDG--PL-----477
Qy 1067 PHADPSAFSAYPGHPPLGLIHTARVLRPPTISNPPPLISSAKAPSVLERQIGAI 1126
Db 478 -----LTPIDGEPMP-----GSDPPADGVRVYGCAGDSREGIWD 510
Qy 1127 QGMSVQLHVPSEHAKAVGVYTMGLPLPMPKTLAP-----PBGVQBOQLS---PRG 1176
Db 511 EDDVRQAARV-----RAAAGVVPVFIPEMGSRQHEALVELIYSGAAGEAMSLQNRPM 566
Qy 1177 QAGPESIGVPTAOEASVLRGTALGSVVGSGITKGIPTRVPSDSAITYRGSITGTAD 1236
Db 567 QA-PQGRNQCQRVYHAPHG--GSFTGSVTPPLPHIG--DAMAAOPLWALPFAVSAV 622
Qy 1237 VLYKGTITRIIGDSPSLDGRG--DSLPGKH--VYEGKKGHVLSYEGMSVTQCS 1290
Db 623 AMSR-----RYDRTOKTFTLIQSIRRAVADMAVYGRAADPAGCATVEAL-CA 668

QY 1075 FSYAPRGHPLGLDHTARPVLP RPPTISNPRLISSAKHPSVLEQIGATISQGMVOLT 1134
DB 2516 KRY-----RLIARFRHEGDEVDVEKP-----DPYQIATTLVYDGLDE-----ISRETDLD-H 2561
QY 1135 VPYSEHAK-----APV-----GVTWGLPLPMDPKLIAPSPGVQO 1169
DB 2562 APIADHANVYHHGDSRIEKNASPVATLEKKOKKVKSPKKEKVDVPAPBEK-----XE 2614
QY 1170 EQLSPRGQAGP-PESLGVPTAOEASVLRGTALGVPG-----GSITKGIPISTRVPSDAIT 1224
DB 2615 IRLIARVTEPEAEIYDVPVAVVETKEKKGDSKSLPTGFFSKTKTKTGTGYPETSE-A 2673
QY 1225 YRG-----SITGTPADVLVYKGTITIIIGEDSPSRIDR-----G 1258
DB 2674 YTGELDTDRJHDELGTSFEHGEHPAYSPKQIVIAI--ETPAPEKKRYLLIARFRHEG 2731
QY 1259 REDSLPK-----GHVIEGKKGHVLSEYGGMSVTQCSK-----BDGRSSSGRPHEETAAP 1307
DB 2732 DEDVVEKDDPYQIATTLVYDGLDEI-SRETDLDPIDHANVYHHGDSRIEKN--ASP 2788
QY 1308 KRTYDMEGRV-----GRAISSASIEGLMGRAIPEPHSPHH-----LKEQHHRGSI 1355
DB 2789 VATPEKKOKKVKSPKKEKVDVPAPBEKKEIRLIARVTEPEAEIYDVPVAVVETKEKKGDS 2848
QY 1356 TQGR-----RSYVEAQEDY-----LREAKULKREGR-----PPPPSRD 1392
DB 2849 KSKLTGFPFSKTYTGTGTPETSEAYTGELDTDRJHDELGTSFEHGEHPAYSPKQI 2908
QY 1393 LTEAVKTOALGPLK-----LKPANEG-----LVATVKEAGRSIHEIPE-ELRH 1435
DB 2909 VIETIETAEPRKRYLLIARFRHEGDEVDVEKPPYQIATTLVYDGLDEI-SRETDL 2967
QY 1436 TP-----ELPAPRLKEGSTTQGTPLKYDTGASTGSKKDVNSL 1476
DB 2968 TPIDAHANVYHHGDSRIEKNASPVATLEKKOKKVKSPKKEKVDVPAPBEKKEIRLI 3024
QY 1477 IGSPPRTPRVAPVLPVADARALERAQVEESLKSPPGTASSGSGSIANGAPVIVELKGP 1536
DB 3025 -ARVTEPEAEIYDVPV-----VKETKEKKGDSK-----LPTGFP 3061
QY 1537 ROSPLVEDHGAPFAGHLPKSGPVTMRRETPRLQEGSISSKASODKRLTSP--REIAK 1594
DB 3062 SKTTTITTTGYP-----ETSEAYTGELDTDRJHDELGTSFEHGEHPAY 3106
QY 1595 SPHSTVPEHHHPISPYEHLKGVG-----GVDLY----- 1624
DB 3107 SPKQIVIAIETPAPEKKRYLLIARFRHEGDEVDVEKPPYQIATTLVYDGLDEISRET 3166
QY 1625 -RSHIPLAFDPTISRGIRPLDAAAY-----LPRHLAPNTYVHLVPRYLIRGVPDTA 1677
DB 3167 DIDHTPIA-----DHANVYHHGDSRIEKNASP-----VA 3196
QY 1678 ALENQTIINDYITISQMHNTATMAQPADMLRGLSPRESSIALNYAAGPRG-IIDL 1736
DB 3197 TLEKKOKKVK-----SPKKEKVDVPAPBEKKEIRLIARVTEPEAEIYDVP 3243
QY 1737 VPH-----LPVL-----VPPETGPATMDRLAYLPTAPQF-----SSRH 1772
DB 3244 VETKEKKGDSKSLPTGFFSKTKTKTGTGYPETSEAYTGELDTDRJHDELGTSFEH 3303
QY 1773 SSSPLSPGPTLTPRTTSSSR-----EBDRD-----R 1802
DB 3304 GEHPFEASSTISITLPAFDPSEHEKVIKPSRFHMFMRMHEDDEVVEKPGYILISSELY 3363
QY 1803 BRDRREREKSLITTTVEHAPIRWPGTEOGSGSGSGSGSGSSSR----- 1850
DB 3364 BGPYNNMPPIGLELPLIDHSQVH-----NGSGKLEKXSSSEKRIKTPSPRVERPS 3417
QY 1851 -----PASHSH-----AHQSPISPTQ-----DALQOR-----PSVLANTGM 1883
DB 3418 EIKLVARIQPLEHVEDSEISEKLSPVKORSRAPSPFMRHOSKORGYPEISPLYEGNLDV 3477

QY 1884 KGIITAVESKPTVLRSTSTSPVRPATPPTHCP----- 1920
DB 3478 TGRAAVEP-NTPTEIVRSVPEHNPQOVSAQBPSSQPNRFNRLIARFRHEGDEDEIEKPD 3536
QY 1921 -----LGGTLDGVYPTLMEPVLLPKKARVARPERPRAD-----THAFLA---K 1962
DB 3537 AYKFTEVYEGRLDIRPV-----VELAQPIIDHSQVYHHGDSMAEKNIK 3583
QY 1963 PRASGLEPASPSPKSGSPRLVPVPGSHATIACTPAKNLARPHNASPP-----PAP--- 2015
DB 3584 PP-----KPVNTKKGLSPEPIESEVHEIRLLTVASSSEEPEDVPIASTVPKPPRL 3638
QY 2016 -----ASADPHEKTKOSKP-----FSIQLELRSLGY-HGS-SYSP----- 2050
DB 3639 RFLSIGKSPSPSPSEPECPQYDGPVDTLSRDELEVMYLLGSPVPPKTPPVKKV 3698
QY 2051 -----EGVEPV-----SPVSPSLT-----HDKG 2069
DB 3699 DKADEERKTYIVIRFRHEGDEPDADKPEAYGFSDDVYTGSLNEISRSELEHAALHDS 3758
QY 2070 LPKH-----LELDKSHLEBELRPKO----- 2090
DB 3759 QVYHDLGSLYKIDNKOKSKDRTELEPKTSRLVARVLPQOASUSGSPKDKDKGFRFR 3818
QY 2091 -----PGPVKLGGEAHLPHL-----RPLPSQSPSSPLQIAPGVK 2127
DB 3819 FFSKASPSKSGYREVSPFPEBPIDVTGRVSELVHPLOSREFPTYSPPKS---QEKPSK 3875
QY 2128 GHQVWTLAQH--ISEVITDQYTHHPQOLSAAPLAPLYSPGAS-----CEVLDR 2177
DB 3876 SSERTEBPVKYHLIARI-----RHEGDEETVENP-DTYGASTSYDGPLETETSKSVLVE 3928
QY 2178 RPPSDLYLPDPDHGAPARGSHSGGKSPPNKNTSVLGGEDGIEPSPPE--GMTPEG 2235
DB 3929 ETPISVSNVYVHHESIRIVERKORAKISAPPEK-----DDQENVPKSPSPFLRG 3981
QY 2236 HSRSAVYPLLYRDEQOTESPMSKSPGNTQPPAFSFKLTES-----NSAMVSKKOBIN 2291
DB 3982 QKVLYSTPSTLYSG-----PLETTDRDLDENILPLSTPYHSSSQHAKRYLI 4029
QY 2292 KKLTHNNEPEYINISQGTIEFMNPAITGTGLM--TYRQAYOE-----HASTN 2339
DB 4030 ARIR-HAGDEDEVAHKLHPFVGFPAQVYGILEKTALCKEVDVPIGRFARVYHSGSS 4088
QY 2340 MGLEAIRKALMGKIDQWEEBSPPLSANAFNLNASASLPAAMPITADGRSD----- 2391
DB 4089 GGLIWNLLKNNNSPASKMDPPKRRKQDKQDKIKLIARVLPKKTPEKSDREKATKE 4148
QY 2392 -----HTLTPGGGGAQVSGRPSSRAKSPAPGLAGDRPPSVSVHSEGD 2439
DB 4149 EPLREAEIYRSTQSEAGNISTVGIYPSLTYDVP-----RPARIEGITYGGDC 4197

RESULT 93
T20774
hypothetical protein ZK270.2d - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20774; T27818
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19322
A:Accession: T20774
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4667 <WLL>
A:Cross-references: EMBL:Z81499; PIDN: CAB54224.1; GSPDB: GNO0019; CESP: ZK270.2d
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20423
A:Accession: T27818
A:Status: preliminary; translated from GB/EMBL/DBJ

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <WMC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.4%; Score 323.5; DB 2; Length 1497;

Best Local Similarity 20.5%; Pred. No. 0.0029;

Matches 379; Conservative 138; Mismatches 646; Indels 683; Gaps 95;

```

QY 716 ENVEAEALHAGNENVRREGSGPATVNNSSDT-----ESIPS 753
DB 71 DKIECPVLCANPITTPREGCCPVCPQTGGDTSPGKGGKGGKGERGLVAVVTGIRGRG 130
QY 754 PHTAEAKD--TGONGBK---PPATIGADGP-----PGPPPPRTSR----- 791
DB 131 PGGPGSGGPRDRGRKGRPRGRGPGIDGEGMPCGAPGAPGRPGHSHPGDMSRPS 190
QY 792 -----APLEPTPASEATGAPPPPPAPSPSPAPPPVPE 825
DB 191 AQMAGLDEKSGIGSVGLMPGSGVGPVGPGRPVGLQGQGGAGPAGPGEPEGPGPIG 250
QY 826 EKEEBETAAPRVYEEGEOKPPRAAEHLAVDTGAEE-----VKSECTEAAE 872
DB 251 SSGPGPPGPKPEBDEG---PGRNGTGVGFSGSGPAGPFGAPGLPGLKGRHKGLE 306
QY 873 GPAKGDAAEAATAEGALKAKEKGGSGRATTAKSGGAPQDSDSATCSADEVDEAEG 932
DB 307 GF-KG-----ETGARGA-----KEAGPTGPGA----- 329
QY 933 DKNRLSPRSLTPTGDP--RANSPQKPLDKOLKORAAI PPIQVTKVEPPREDAA 990
DB 330 -----MGF---LGRPMGGERGRLLPQ-----GAPKRGGA 356
QY 991 PTKRAPRAPPPQNLQPEDAPQOR-----SSPRGKSRAPPADKEAFA 1037
DB 357 HCMPGKPGMGLGIPGSSGFPGNPMKGERPHGARPEGPGQGETGPGC---PA 411
QY 1038 EAQKLPGRDPCWTSGLPFVPPREVIKASPHAPDPAFSYAPPGHPLPLGLDHTARVLP 1097
DB 412 GSQGLPG--AVGTG---TPGRKATGSAAGTSGPGLA--GPPSGSPGQ----- 454
QY 1098 RPPITSNPPLISSAKHPSV--LEKQIGAI SQ-----GMSVOLHVPYSEHAKAPVG--PVTM 1150
DB 455 -----STGPGIRGQSGDPVPGPKGBAPKGEPRPHIGQIPGPEEGKRGRPDGPTV 510
QY 1151 GLPLPDPKCLAPFSVCQKQEOLSPRQAPRPESLQPTQOASVLKGLTGLSVPGSITK 1210
DB 511 GPPGPMG-----ERGAP--GNRGPFSDDGLPGRGAQGERGVPSSGPKG--Q 555
QY 1211 GIPSTRVPDSALITYGSIITHTGTPADVLYKGTITRI--IGEDS---PRLD--RGRDEL 1263
DB 556 GPPGR--PGEPLPGAKGLT--GNPGVQGRBKGLPLGARGEDGRPPPSIGIRQPGSM 612
QY 1264 ----PKGHVYEGKKGHVLSYEGMSVTQCSKEDGRSSGPHETAAPRTYDMMEGRVG 1319
DB 613 GVPGRKSSGDLGKPG-----BAG-----NAGVQGRGAPGK-----DGEVG 649
QY 1320 RA--ISSASIEGLMERAI PPRHSHNLKEOHNRKISITQGI PRSVYEAQEDYLREACT 1377
DB 650 PGGVPGRPLAGEBGAAPP---GPTGF-----QGLP----- 678
QY 1378 LKREGTPPPPPSR-----DLTEAYKTQALGPL-----KLKPAHEGLVATVKEAGRSIHE 1427
DB 679 -----GPPGPPGEGGKADQGVPRGPAVPLGPPRGENGNGERG-----EPG--ITG 724
QY 1428 IPREELR--HTPELPLADRPLEKESI-----TGSTPLKTDYTGASTGSKKHIVSLI 1477
DB 725 LPEGKEMAGGHGDDGKG--NPGTGTITGTPGGLQGMF--GBRGIAAGPGRKGD--RGSI 780
QY 1478 GSGPRFFPVNPLDWAADARALERACYESLSKSRPTASSGGSISAKGAPVIVPELKGK- 1536
DB 781 GSKG-----AEGTAGNNGARQLPPLAPPPGAGLILGAP-----GEPG 817
QY 1537 -----RQSLTYEDHG-----APFGHLPRGS---PVTMRBPTPLQSGSLSSSKA 1579

```

```

DB 818 PRGLVPGPSRGNPCPSRGENGPTGAVFAG--PQSDQPGVKSGPGEFGQGDAGS--- 872
QY 1580 SODRKLSTPREIASPSTVPEHNPPIISPENHLRREVSGVDLYRSHIPLAFD--PTSTP 1638
DB 873 -----RGPQGLASFP-----GPHGPHGV--PGLKSGRGTGQF-----PGATGPGSGAG 913
QY 1639 RGIPLDAAAYVLPRLHLPNPPTVPHLYP-----PVILRGVPTDALENROTIIINDYITSQ 1694
DB 914 RVGPPGPAQ-----APGAPGAPGEPKKEPPLKREDPGS----- 947
QY 1695 MHNNTATAMAQRADMLRGLSPRESSLALNYAAGPGIIDLQVPHLPVLVPTPGTPATA 1754
DB 948 -HGVVD-----RG-----PAGP-----PGSPGDK 966
QY 1755 MDRLAVLPAP-----QPFSSRHSSPPLSPGCTHTLTKPTTSSSRERDRDRERDRR 1810
DB 967 GD-----PGEEDQP-----GPDGPPGAGTTGORGIVGMDQR----- 999
QY 1811 EKSLTSTTVVHAPIMWPGTEQSSGSGSGGGG--GSSSRPASHAHQHSPISPRTQ 1868
DB 1000 -----GYTMGGLP--GRAGTGKVGPTGATGDKPPRGVGRPGSNGPVGEGPBPAGN 1052
QY 1869 DALQQRPSVLTNTGKGIITAVEPSKPTVLRSTSSPVRRPATPPATHCPLGCTLDGV 1928
DB 1053 DGTPRDGAVERGDRG---DPC--PAGL-----PQSGCAPGPGVGAAPDAG 1096
QY 1929 Y---PTLMEPVLLPKEAR--VARPERPADTGHAFLAKPRASGLERASSSKSEPRP 1983
DB 1097 QRGEPGSRPVPGRPAGKRGILPGQGRPGDKG-----DNDRDRDQD- 1139
QY 1984 LVPPVSGATTARTPAKULAPRHASDPDPAP--PASASDPREKTQSKPFSIQELETRS 2040
DB 1140 -----KCH-----RGTGLQGLPGRPPGMBGSGAGIIPPFRGPRP----- 1176
QY 2041 LGTHSSYSPEGEVBPVSPVPSPLTHDKGLPRHLEBLSKSHLEGILRPKQRPVYLGEA 2100
DB 1177 -GPVSSSGKEGNRPLGPIGPPGVNAGSVG-----EAGREGP--PDEPBPDPGPP 1224
QY 2101 AHP-----HL-----RPLP---ESQPSPLQLTAPGVKQGRVVTLAQHSVIT 2144
DB 1225 GHITLALGDIMGHYDENMDDPLPEFTEDQAAADDNTKDPGI--HYTLKSLSSQLETMS 1282
QY 2145 ODYTHHPQOLGAPLPAPLYSFGASCPVLDLRPPSDLYLPPDHGAPARSGPSEGG 2204
DB 1283 PDSKKHPR-----TCDDLKCHT-----KQSGE 1308
QY 2205 RSPBNKTSVLGGEDGIE-----IVSPE--GMTPEGHSRAVPLY 2246
DB 1309 WYIDNQ-----GSADDAIKVYCNMETGETCISANPASVPRKTMWASKSPDNKPVWYGLDM 1364
QY 2247 RQGEQTEPSRMKSAPGNTSOPPAFFSKLTENS-----AMVSKKQELINKLKN 2295
DB 1365 NRGQF--TYGDYQSPNTAITQMTFFRLSKASNLTYICNTVGYMDQAKMLKAVV 1422
QY 2296 THNRMEPEYNIQPGTEIFNMPAITGTGLMTRSOAVOEHASTNMG 2341
DB 1423 LKGSNDLE-----IKGBGNIRRYTVLQDTCKRNG 1453

```

RESULT 91

S23810

collagen alpha 1 (XVI) chain precursor - human

W:Alternate names: procollagen alpha 1(XVI) chain

C:Species: Homo sapiens (man)

C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 15-Sep-2003

C/Accession: S23810; P00612; S08012

R:Pan. T.C.; Zhan. R.Z.; Mattei. M.G.; Timpl. R.; Chu. M.L.

proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992

A>Title: Cloning and chromosomal location of human alpha1 (XVI) collagen.

A/Reference number: S23810; MUID:9235339; PMID:1631157

A/Accession: S23810

A/Molecule type: mRNA

QY 1562 MREPTRLQSGSLSSKASQDRKLTSTPREIAKSPHSHTVPBHHHPISPEHLLRGVGV 1621
D 1335 -----RTR-----TTSKPSQSIAKEKGDALSLNLT-----EQHQPRLDAQEELKEISEQ 1381
QY 1622 DLYRSHIPLAFDPTSPRGIPRLDAAAYLLPRHILANPTYPHLYPRLIRGYPDTALEN 1681
D 1382 D-----GSDQSF-RTPRMD-----LKSQDEBEL-----ESLGTSSANQD 1416
QY 1682 RQTIIND-----YITSQMHNTATATMAQRAIMRLGLSP-----RESSLALNTAAG 1727
D 1417 CPQGVKPRMOMOVQLRRQNHSLIEDLRRLDQAEHLLENODGVERRDRIRIGIATLSAS 1476
QY 1728 PRGIIDLSQVHPVLVPEPTGTPTATMDRLAYLPTAQPPSSRRHSSPLSP----- 1779
D 1477 HQDCQTGSEEPQVSGVPESP-----AATER-----OEPLSPQSGDPVGPRAIKKGL 1525
QY 1780 ---GGPHTLTKPTTSSSEERERD--DREDRDREREKSIITSTTV-EHAPIWRPCTEQ 1833
D 1526 SRQGRREGQKPLRGFQRSGADPQDVNVGSHVAKSKILDQQRLLQDHGIMKPSREQ 1585
QY 1834 SSGSS-----GSSGGGG-----SSSPA--SHSHAHQHSPISTPTQAL 1871
D 1586 PSSASDLPCSSDPSLVNDRSIRDSGPLKLPISGRPSFRPYRRPQERSQGSHEAQAQ 1645
QY 1872 QQRPSVLHNTGKGIITAVEPSKPTVLKSTSTSPVPAATPPATHCPLGGLDGVYPT 1931
D 1646 RRCESADHHPDDSLIIDL-----IDTSNRSPMPMP----- 1677
QY 1932 LMEPVLPKEAPRVARPPRPRADTGAFLANP-PARSGLEPASPSKSEP 1981
D 1678 -----LPEEKPMALQALHQLKLAAYHAFRCLEPTAISRLP--RRQNTPEP 1720

RESULT 88

T48818
glucan 1,4-alpha-glucosidase related protein [imported] - Neurospora crassa
Nucleotide names: protein 68B2.20
C:Species: Neurospora crassa
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48818
R:Schulte, U.; Altm, V.; Hobeisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224541
A:Accession: T48818
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2022 <SCH>
A:Cross-references: EMBL:ALJ53821; GSPDB:GN00112; NCSP:68B2.20
A:Experimental source: cosmid contig 68B2; strain 74
C:Genetics:
A:Gene: NCSP:68B2.20
A:Map position: 2
A:introns: 1192/3

Query Match 2.5%; Score 324.5; DB 2; length 2022;

Best Local Similarity 18.7%; Pred. No. 0.00037;

Matches 430; Conservative 274; Mismatches 892; Indels 701; Gaps 102;

QY 533 EEKPEVENDKEDLLKTKTDDTSGEDNDEKAV--ASKGRKTANSQGRKRITSMANE 589
D 62 KKKPELKKPEPELT---GDFLLGRPSQPVAAQASAKRKT-----MGMMD 104
QY 550 ANSEALITPQGSALASHELNSSRWTEBEMETAKGLLEHGRMSATAAM----- 640
D 105 AONVEELRQERAAAVAKLQOPG-----GVADRVAQOKASQAQAAVKAELPT 152
QY 641 -VGSXTVQCKNFYNYKKRQNLDEILOHKLME--KERNAPRKKRKAAPAAASEAAPP 698
D 153 FVAEABAREPTFAANLGS---DDVDEDRKIRTKRQKPPKQKPKT----- 196
QY 699 VVEDEMAAGVSGNEEWEAEALHASGNEVPRGEGSGPATVNNSSDTESIP----- 752
D 197 -----ETVEPKGTSGSDSTAKEKNKVLGRMADRP-----DIVIKEARPPGLPKKRIIS 246

QY 753 SPHTAAKDQNGCPKPPATLGADGPPPPPTP---RTSRAP-IETPASBATGATP 808
D 247 DDHMKKQKQKGLAKSPV-----PEAGQPIPDFLKRTAQNSVQKIRDMQRCLEP 300
QY 809 PPAPSPSAPPVVPKKEKEETAAPRVE---GEOKPPAAEELAVDTGAEEPVK 863
D 301 PPPPVARRPTVTKYRAKGTETVVEDEDDHNRGLSPDRPRKRSSSGSSRRSS 360
QY 864 SECTEA-----EEG-----PAKQAAEAATBEGALKAEKKEGG 899
D 361 NPRTQCCPNDGIVRSILOKKSNDGIRISPVPARSLPDGQIKVGPVVSADSSRS 420
QY 900 SGRATTKSSG-APQDSASATCSADYDEAGDKNRLSPR-SLTPPGDRAVAP 957
D 421 ISTVSPSSSGRTSDSGSKT-----PPGASPPRRASTP--PRKASTP 464
QY 958 QKPLDLKOLKOR---AAATPIQVTKNHPRRDAAPTKBAP-----APPPONLQ 1007
D 465 LRKASTKPRARSDHSAASDDVIEV--IVESEVSXKSPSPPKRLRSPPPKRLP 522
QY 1008 ESDAPQPGSSPRGKSPAPPADKEAPAAEQKLPQDPCTWGLPPPV-PRVVKAS 1066
D 523 RRRG-SRQGARKKRKRSPSPPT---ATQETTTDDR--RGADKPMPTPRNNGSS 574
QY 1067 PHADPSAFSVAAPPCHPLPLGLHDTARVLPRLPTISNPPLISSAKHPSVLERQIAIS 1126
D 575 GEDSD-----RRPTAAGIDL-----AEIP 595
QY 1127 QGMSV-QLHVP-----YSEHAKAPGVPTWGLPLPM--DPKCLAP-----FGVKQEQ 1171
D 596 FGSAFSELEPLAGHTQRTAQAQAKP---KKPQONESLKVLPVNLKKVMTGA-MEK 650
QY 1172 LSPRQAGPPELSLVPTAQEASVLRGTALSGVPRGSIITKGIPTSRVSDSAITRGSIIT 1231
D 651 MQEAAEPPRPPTKQKPARIESMNLNTVDPVEG--MPNLPVVAEPRLVS----- 700
QY 1232 GTPADVLYKGTITLIGEDSPSLDRGDESLPGHVIYEGKKGHVLSYEGMSVTOCSK 1291
D 701 -TPE--RNSKEKLVDRDLPAHREBAPRN-----SK 728
QY 1292 EDGRSSSGPPHETAPKRTYDMEGRVGARAISASIBGLMGRAIPREHSPHILKEQHNI 1351
D 729 EKLVDITVSTHKATPE-----ISKXASAELEPRKXRPQPVVHEELRKST 775
QY 1352 RGSITQGIPRSVEAQEDYLRE-----AKLLKEGTPPP-----PSPRD 1392
D 776 RSDSEVTPK-----RRKSPTTTPDSAGLKRSTRSPTLPRRSSGSVKKPFRD 825
QY 1393 -LTAAYTQNALGPIKINP-AHEGLVATYKAGRSIHETPRELHTPEPLAPRLKGS 1450
D 826 VLKEAFKGS-SANHLAPMVPSCETDVESEPETEHD--ESRRSPQOQSPDSYKRS 881
QY 1451 ITQGTPLKYDTGASTGSKKHVDVSLIGSPKRTPPPVPLDVMDARALERACYEESLKS 1510
D 882 ASPRPSRADSYSSSEPTR-----GPRRRRPTSDLDLSTIISEDREKQEPYXK 933
QY 1511 RPTGASSGSGIANGAPVIVELKPKROSPLITYEDHGAPFGHLPKSGPVIMREPTP-- 1568
D 934 KD---TDSVSTVSGSTVQAPEBAFGOPTPLTREASQOSQTSIRKSGSLKRLITTHS 990
QY 1569 -----LOEGSISSKASQDRKLT-STPREIAKSPHSV-----DEHHPHPISP 1610
D 991 DLVSVLSLPDGGQLVPPERSSTIASRSLHKKPSKANDSRVNDLLEFADENHFYH--RE 1048
QY 1611 YEHLRGVGV-----DLYRSHIPLAFDPTSPRGIPLDAAAYV--LPRH-- 1654
D 1049 LKTIYDGVVPLVPLNEFVHGDVADARKTDSMA--KAVVNVGVALKEKLMYHKAAPLHDI 1106
QY 1655 -----LAPNTYPLHLYPRLIRGYPTALLENKQTIINDYITISQMHNTATATMAQRA 1707
D 1107 RRLLEWLEAVSPVY-----NNYLDVWRLGQDILVNLAPSPGKIDENDS----- 1150

QY 1303 ETAAKRTYDMMEGRVGRALSSISIEGLMGRALPERHSPHILKEQHILRGSTIQGIPRS 1362
DB 1101 GTPGP-----AGLATTLP-----LTTPRAQGRAPALSSS 1133
QY 1363 YVEAOEDYLREAKLKEGTTPPPPSRDLTEAYKTQALGPLKLPAREGLVATVKEAG 1422
DB 1134 WQPPAN--MNRPEPSCRDTAPPTMLISQSPA---EADGSVAVPQAGVARIPIEPR 1188
QY 1423 RSIEHIFREELRHTPELPLAPPLKSGSTQGTPLKYDTGASTTSSKKHIVSLIGSGR 1482
DB 1189 TSSHADPE-----ABPPWSGRLPAFGVLPAT-----EPRGTPGSPSG 1227
QY 1483 TEPVPHPLDVM-----ADRAL-----ERACYEESLKSPGTASSS---G 1519
DB 1228 TQEPGGLGLEKPLRQPPPEKALDLEKPLRQPPPEKALDGLISQSGALATQOMLG 1287
QY 1520 GSIAAGAPVIVPELGE--PROSP 1540
DB 1288 GQ--RG--VRVPLLSRLPYQPP 1306

RESULT 87

T26135
hypothetical protein Y6B3B.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 28-Oct-1999
C:Accession: T26135; T27328
R:Doonan, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20158
A:Accession: T26135
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <W12>
A:Cross-references: EMBL:Z82068; PIDN:CAB04901.1; GSPDB:GN00019; CESP:Y6B3B.1
A:Experimental source: clone W04A4
R:White, S.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20345
A:Accession: T27328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <W12>
A:Cross-references: EMBL:AL032655; PIDN:CAA21727.1; GSPDB:GN00019; CESP:Y6B3B.1
A:Experimental source: clone Y6B3B
C:Genetics:
A:Gene: CESP:Y6B3B.1
A:Map position: 1
A:introns: 20/2; 53/3; 82/3; 329/3; 332/3; 355/3; 396/3; 462/3; 1799/3

Query Match 2.5%; Score 325; DB 2; Length 1829;
Best Local Similarity 19.2%; Pred. No. 0.00032;
Matches 356; Conservative 221; Mismatches 616; Indels 659; Gaps 88;

QY 370 HEVSEIIDGLSECEENLEKQROLAVIPMLYDADOORIKFI-----NANGLMADPM 420
DB 289 HEIAGLP--WAOEALSKLERG-----YDA--RIHEIMAGNSAIDQASNDIQAYGA 339
QY 421 KYVKDQVNNMWSSEOKETFRKKFMQHPKPNFGLISFLERKVAECVLYYYLTKKENYK 480
DB 340 KI-----QILE--EPVEDQTSKKVLLQDPYK-----DQKFGESVQOVVLEKIGESKK 385
QY 481 SILVRSYRRGKSQQQQQQQQQQQQQQQQPMP--RSSQEKDEKEKEKAEE--EKEKE 537
DB 386 -----DQKKIEDVQLQENLELEFPGRPAVEEDPKQVVEGISKNVEEDE 432
QY 538 VE--NDKEDL-----LKEKTDGSGNDKEKAAVASGRKTANSQGRKRITRSM-- 586
DB 433 VELDHEDLTATPLSHMDVTENIONLDANAAEPIENPLAAVDEQIKADKILAEVQI 492
QY 587 -----ANEANS-----EAATP--QQSABL-----ASMEINSS 613
DB 493 QDDMHQNPALSESDIMEAREAQTVQDFIEDPVTEVEVETELPEDRKDPALHELENN 552

QY 614 RWTEEMETAKKGLLEHGRNMSALAPMVGSKTVGCKNFYFNKKQQLDELLOOHKLK- 672
DB 553 DPTADVDPSVDVDFND-----LVDS-----DEALEVELKD 584
QY 673 ---MEKENARRKKKKAPAAASEAAPPVVEDEMEASGVSGNEEVEAEALHASGN 729
DB 585 QKVSQETSADQCK-----SEADKDLGSEGSBDQAPELGKGDESESKDPEVABAN 637
QY 730 EVPRGEGSPATVNNNSDTESIPEPHTEAADTQONGKPPATIGA--DGPPG-----PP 783
DB 638 LIP-----SVPRSGDLQATDSKESHSDXES-----KDPQVLDANNDPQSGESKDP 684
QY 784 TP-----PRTS-----RAPLETPASEATGATPPAPSPSAPP 819
DB 685 SLESKDHQVLKVLIPREGSVESKNHOQAAGVHPHAGSESKGTEDPNTVLESPPAPK 744
QY 820 -PVVPKEKEEET-----AAPPV--EKGEEKPPAA----- 848
DB 745 NPVEHNLEAQNPDPPIADPAPPIRIKRYKVKVPAVGVYKTRTQKMKLEDRD 804
QY 849 ---BELAVDTGKAEPEPVKSECTEAEF-----GP-----AKGDAEAAETEGALK 892
DB 805 RTSMEEVAAKTEKQDPASBDLSDBRELSVAAGPQAPKRRRRGRMEDEE-TARKRRK 863
QY 893 AEKKGSGGATTAKSGAPQSDSASATCSADEVDEAGGDKNRLSPRPSILPTGDP 952
DB 864 GRKLE-----ASEDPDQSGAPGAPPEPVKKYKTSRKRYVDSENV-----PQ 907
QY 953 ANASPOKPLDLKOLKORAAAIPPIQVTVKHEPPREDAAPTRAPAPPPQNLQESDAP 1012
DB 908 MNTRRK---ISEREYPAELGVPYRRKRKDDPEVDAQ-----AQ 944
QY 1013 QQPGSSPR-----GKRSRAPPADKAFAPAEAKL-----PGDPPCWT 1051
DB 945 KDQGSATWVCIKIRFDGIPKIRDPA-----KSGAEDDELNFTSSDVLQCGSGVCS 999
QY 1052 GLPFPVPP-----REVIKASPHAPDPSA-----FSVAPPGLPLGLHDTARPVLP 1097
DB 1000 ---PATRSEBQDSTQSGEDSDTSKIPCFEVSMP-----ISEPII- 1042
QY 1098 RPTTISNPPPLISSAKHS-----VLEROIQAI--SQGMSVOLHVPYSEHAKAPV-G 1146
DB 1043 -EALIKNDPILIFNAGSGSSASSEAGPRIGIYEDQFOUMIGNIPGDDSGSVQ 1101
QY 1147 PVTWGLPLPMPBKLAFFSGVKGQEL---SPRQAGPESLGVPTAGASVAGTALGS 1202
DB 1102 SETNKDFEPMKP-ELQDVAGTDQTRILLATSPDSEA---SANDHAKDPNT--ETTKK 1154
QY 1203 VPGSITKGIPTSTRVPSDAITVYRGSITHTGPADVLVYKGTITRILIGEDSPRLDRGRDS 1262
DB 1155 LEAEKPOQOFKVVYFKQKLE-----SEEDGNGKT-----DPOQDL 1193
QY 1263 LPKGHVIEGKKGHVLSYEGGMSVTQCSKEDGRSSSGPHETAPKRTYDMMEGRVGRAL 1322
DB 1194 M-----LYKSGREHIDL-----ELIOEKASYQSENNM--EI 1225
QY 1323 SSASIEGLMGRALPERHSPHILKEQHILRGSTIQGIPRIVEAOEDYLREAKLKE 1381
DB 1226 AOPSVSSEMSSDLPA--GPHHVLKEORRDTSSFSF-----AAYLADLKKIYKN 1271
QY 1382 GTPPPPSRDLTEAYKTQALGPLKLPAREGLVATVKEAGRSIHEIPREELRHTPELPL 1441
DB 1272 ISIAEQSGSDSTN-----PNSDKDC----- 1294
QY 1442 APRPLKESITQGTPLKYDTGASTTSSKKHIVSLIGSPGTFPPVPHPLDVMAARALER 1501
DB 1295 -----RKREL-----LPQLDRSQIA----- 1311
QY 1502 ACYESLSRPGPTASSGSGSIARGAAPVIVPELGRKROSPLYEDHGAFAAGLPRGSPVT 1561
DB 1312 ---REDLK-----PKISTPEES---ED-----PARSOVT 1334

QY 2029 KPFTSGLSLGSLGSSSYSGVEPVSPVSSPLTHDKGLPKHLELDKSHLEGELRP 2088
DB 1029 -----GFTSTGAP--GPIGPTGARGST-----GPAKP 1054
QY 2089 KQP-CPVILGSE---AAILPH--LRPLESQPSSSPLQTAQVKGHQRVVTLAQHISEV 2142
DB 1055 SGPSPPAEBERGRTGTAGKHGHPGVSGLGLOCTSGPMGE--PGARGQ----- 1100
QY 2143 ITQDYTRHHPOQLSAPLPAPLYSFPGASCPLYDLRRPPSPLYLPPDHGAPARQSPHSEG 2202
DB 1101 -GQOQSTR-----GLPGA-----RGSNGNDPSSGPR 1124
QY 2203 GKRSPE-NKTSVILGSGDGIETPVSPPEGMTEPGHRSAGAVVPLLYRDEQTEPS-RMGSK 2260
DB 1125 GPDPGEGRGPRGSGSSGPPGPPGPPGPPGQVQSS-YGVR-----PSFGSGK 1176
QY 2261 SPGNTSQPPAFPSKLTESNSAMVKSQKQEIINKKLTHNRNREPVNISQPTGTEIFMMPA1T 2320
DB 1177 GQGS--PYGAYRDDSKNDVAKIQDTL-----LGAIS 1208
QY 2321 GTGLMTYRSQAVQEHASTN 2339
DB 1209 ALGQQIELIKAPQCKAKTN 1227
RESULT 86
T09219
basal transcription factor SNAPc large chain SNAP190 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09219
R:Wong, M.W.; Henry, R.W.; Ma, B.; Kobayashi, R.; Klages, N.; Matchias, P.; Strubin, M.;
Mol. Cell. Biol. 18, 368-377, 1998
A:Title: The large subunit of basal transcription factor SNAPc is a Myb domain protein b
A:Reference number: 216616; MUID:98078693; PMID:9418884
A:Accession: T09219
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residue: 1-1469 <WON>
A:Cross-references: EMBL:AF032387; NID:G2641556; PIDN:AAC02972.1; PID:G2641557
A:Experimental source: tissue type fetal cell teratocarcinoma
C:Genetics:
A:Gene: SNAP190
C:Function:
A:Description: transcription factor; required for transcription of snRNA genes
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo
C:Keywords: DNA binding; RNA biosynthesis; transcription factor
F:396-447/Domain: myb DNA-binding repeat homology <MYB>
Query Match 2.5%; Score 326.5; DB 2; Length 1469;
Best Local Similarity 20.1%; Pred. No. 0.00022;
Matches 331; Conservative 196; Mismatches 573; Indels 543; Gaps 76;
QY 84 LRPEHSYLPETLKGEMEPTEISKRPRLLELPPLLRPS-PLLAT-----GQPSSEDLTKD 138
DB 21 LDPGSGSHVEISSESL-----SDEABDLPSEDLPADPPISEERMGESANDEDKDD 76
QY 139 RSLTGKLEBPVSPSPHPTDPELELVPPRLSKEELIQNMDRVDRITMVEQOISKLKKQO 198
DB 77 KTL-----PEDPETCLQNMVYQVIEQKLA-----ANLLAQNRQGE 116
QY 199 QLEENAA--KPPPEKPVSPPIESKHSLSVOITIDENRKAEAARHLEIGIPOVELPL 256
DB 117 ELMRLASCKTKVDGKSLPSTYMGHMKPYFKDK-----VTGVPAN-- 162
QY 257 YNQPSTDRQ-YHENIKIKINQAMRKLLIVKRNHA--RKQKQKFCORYDQ-----LMEAL 309
DB 163 -----EDTEKKAQGIK--AFELLVTKKMKMEKALKRSVSDRLQRLQKLLKLEVL 215
QY 310 EKKVERIENPRRAKESKVRREYKQPEIRKQRE--LQRMQSRVQSGSGLSMSNA 366
DB 216 HQKSKVSELSERQMLEKQGR-AEKELIQDINQLPEALLGNRLDSHDEKISINIFES 274

QY 367 RSEHEVEIIDGSEQENLEKQMLAVIPMLVDADQQRKIFINMGMLMADPMKVYKDR 426
DB 275 RSAELIRKFMON-SEHSINKQ----- 295
QY 427 QVNMNMSEOEKETREKFMQHPKNFGLIASLEKRTVAECV-----LYVLTKNENYKS 481
DB 296 -----EWRREBERLQALAAH-----GHLEMQKIAELIGTSRSAFOCLQCKQOHKA 343
QY 482 LVRSYRRRKGSGQQQQQQQQQQQQQQQPMRBS-----QEKDE-----KEKEKA 529
DB 344 LKRREW-----TEEDRMULTQLVQEMRVGSHIPYRIIYWEGRDSMOLIVMTSLDGL 399
QY 530 EKEKEKEVEVNDKEDLLKEKTDPTSGDNDKEKVAASGKRTAASGRRKGRITRSMANE 589
DB 400 KKGWAPR-----EDAKILOVAKYGEODWKITEEVPGRSDACQDRYLRLHPSL---- 451
QY 590 ANSEBAITPQOASAEIASMELNBSRWTEEMETAKGLLEHG-RNWSAIARVGSKTIVQ 648
DB 452 -----KGRMMLKEBEQLIELIKYGVGHMAKIASELPHRGSQ 490
QY 649 CKNFYFNKKQNLDELLOHKLMEKERNARRKKKAPAAASEBAAPPVVEBEWEAS 708
DB 491 C-----LSKWKIMGKKQGLRRRRRRARRS-----VRWSTSS 524
QY 709 GVSGNBEEMVEAEALHASGNEVPRGEGSPATVNSSDTSTSPHTEAKDQNGQPK 768
DB 525 GSSS-----GSSGSSSSSSSSSEDEP-----EQAQAGEG 555
QY 769 PPATGADGPPPPPP--TPPRTSRAP1EPTPASEATGAPTPPAPSPAPPVVPEKEE 826
DB 556 DRALLSPQYWPVDMDLWPAQSTSGPWR-GGAGMWGGPAPSLSPRGSS--ASQGS 611
QY 827 KEETAAAPVYEGEBQKPPPAEELAVDTGAEEPV-KSCTEAEAGPAAKGDAAEAE 884
DB 612 KEASTTAAAPBEETSPOVPA-----RAHPVPVPSAASASADTRPA-GAEKQALE 661
QY 885 A-----TAGGALKAEEKEGSGRATYAKXS--GAPDSSATCSADEVDEA--- 929
DB 662 GGRLLIYVETTVLRVLRANT-----AARCTQKQLOKRP1PTSPGVSSGSDVARSIVQ 717
QY 930 -----EGGDK-----NRLSPRPSL-LTP-TGD--BRANASPOKPLD----- 963
DB 718 WLRRHATQSGRRMRHMLHRRLRLRLLAATVPWVGQVVCQTAAS-QRAVVGQIADGL 776
QY 964 -KOLKORAA1P-----IQVTYHPPREDA--PTKAPRPPPPQULQPSD 1010
DB 777 REOLOQARLASTPVFTLFTQLFHIDTAGCLEVVERKALPRLPQAGARDPVHLLQASS 836
QY 1011 APQPGSSPRKSRSPAPPAADKFAFAAQAQLPGDPCMTSGLPFP-VPREV1KASPH- 1068
DB 837 SAQS-----TGHLFPVNPADBAKSAASHK 861
QY 1069 -----APDSAFYAPPGHPLPLGLHDTARPVLRPPTISNPPLI----- 1109
DB 862 GSRLASRVRRTLPQASLASTGP-----RPKXTVSE--LLOEKRLQ 903
QY 1110 -----SSAKHPVYERQIGAISQMSVOLHVPYSEHAKAPGPVTMG1PL--PMDPKLA 1162
DB 904 EBARAEATRGVIVLPSC-LVSSSVILQPLPHPHGPAAGPVLVNLPLSGPAPAAK 962
QY 1163 P-----FSGVKQEO1SPRQAGP-----PSIG-VTPA0EASVY-----R 1196
DB 963 PGTSGSOWEAGISADKRLSTM-QALPLAPVPSAEAGTAPPAASQAPALGPOGISVCPES 1021
QY 1197 GTALGVPGGSIITKG1PST-----RVPSDAI1YRG-S1TH-----GTPADVLYKGT 1242
DB 1022 GLGQSGAQAARSKQGLPAPFPPLPGABSPPTLPVQPSLTHIGPHVATSVPLPVTWLT 1081
QY 1243 ITRITIGEDSPRLDRGEDSLPKGHV1YEGKKGHVLSYEGGMSYTCQCKEDGRSSSGPH 1302
DB 1082 AQGLLPVVPVAVV-----SLPR-----PA 1100


```

Db      953 VHSILVAVPFTLGQKELLGRTGLVEEAFWIDKIKSCHFYVSTVEEAVATRTALHGVK 1012
      1475 SLIGSPGRTFPVPLVDMAADARALERACYEES-LKSRPGTASSSGGSIAGAPVIYEL 1533
      1013 -----WPOSNPKFLCADYABODELDYHNGLLVDRSPSTKTEEGQIPPLHPPPPP 1603
      1534 GKPROSPLT--YEDGAPPAAGHLPRGSPVYTWSEPT-----PRLQESLSSSKASQDR 1583
      1064 VPPQGFPAEOREOEBAVAEOWAEERERERERERERERERERERERERERERERERER 1122
      1584 KLTSTPRELAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSLPRGILP 1643
      1123 R-----RKERAKSEKKEKKEKAKOEPPAKLLD-----DLFRK-----TK 1158
      1644 DAAAYVYLRHLAPNTTYPHLYPPIYLRIGYPTALLENQTIINDYITSQQMHNHTATAM 1703
      1159 AAPCIYMLP-----LTDSDIQKEAEBAERAKERERKREKREKREKREKREKREKRE 1207
      1704 AORADMLRGLSPRESSSLATNYAAGPRGIDLSQVPHLPVLVPTGTPTAMDRLAYLPT 1763
      1208 RENNRLERERERERERERERERERERERERERERERERERERERERERERERERER 1228
      1764 APQPFSSRHSSSPLSPGCPHTLTKPTTSSSERERDRDR-ERDRDRERER 1812
      1229 -----ERERERERDRDRDRDRERDR 1249

```

RESULT 85

```

S23809 collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)
C/Species: Strongylocentrotus purpuratus (purple urchin)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A/Accession: S23809
R/Expósito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 267, 15559-15562, 1992
A/Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) col1
A/Reference number: S23809; MUID:92348411; PMID:1639795
A/Accession: S23809
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1414 <EXP>
A/Cross-references: EMBL:M2040; NID:g161435; PID:AAA30035.1; PID:g161436
C/Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F/1207-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

```

Query Match      2.5% Score 326.5; DB 1; Length 1414;
Best Local Similarity 20.2%; Pred. No. 0.00021;
Matches 356; Conservative 110; Mismatches 576; Indels 717; Gaps 85;

      729 NEVPRGCS-----GPAIVNNSDTEISIPSPHTEAKDTQNGKPPATGADGPP----- 779
      38 SEGRGDKGKGGEFGDADINSANPPGLPGP-----VGPSPGSPGSPA--GNNGPDPNG 91
      780 -----PGPPTP-----RRTSRAPTEPT-----PASEATGAPT 808
      92 PRGNPGMDGLTGLRIGIPGPSPGKSGSLVASAQTSSFNKPSLAGYQYQADLAAGTGP 151
      809 --PPAPSPAP-----PVVPEKEKEBETAAPVEGEEOKPPAEELAVDTGKAEPV 862
      152 RGPFGPGRSGPQGLTGTSGSGSGTGPBGNGSGPGLPGRFPGS-----DQDDGTF- 203
      863 KSECTEAEESPAGKDAEAALATAGALKAEKKGSGRATTYKSGAP--QDSDSAT 920
      204 -----GSQQRGPA-----GTPGSRGTGMGAPGMKGHGLPG 237
      921 CSADVDADAGDGNKRLSP-----RPSILPTGP--RANASPOKPLDKOLKORAAAIIP 974
      238 MTGSGEGEGEGEGSGDSGPGVGAFAAGSGQGERGKRGTP-----AGQDGRGA--- 289
      975 PLQVTKVHEPPREDAAPTKPAAPAPPPQNLQPSDAPQDPG--SSPRKSSRP----- 1026

```

```

      290 -----DGATSGQPPRGSGP-----AGAPGMPIGSAKGDAGSPGARGSPG 330
      1027 -----APP-----ADKEAFAAAOQLPGDPPTCWTGGLPPVPVP 1059
      331 LGARGRGSESGQGTGPPGVGRDGSNGAKGSAPSGAQGTGPP--GARGPVP--- 385
      1060 REVITKSPHAPDPSATSVAPPCGHPRLGLHDTARPVLPPPTISNPPPLISAKHPSVLE 1119
      386 -----AGSPGPAGSKDQGNPGQP-----GAQESGFLGPGGETGPAPP----- 425
      1120 RQIGALSQGSVOLHVYSEHAKAPVPGVYTMGLPLPMDEPKLAPFSGVYQEOQLSPGQAG 1179
      426 ---GA---QGSRGSG--SRGALGAPGP-----GGVGE-----RGPMG 456
      1180 PPESLGVPTLOEASVLKCTLGSVPGSGITKGIPTSTRVPSDAITRGSTTHGTPADVLX 1239
      457 PPGMSGAPGA-----PGAKGDRGLPG----- 477
      1240 KGTITRIIGEDSPERLDRGREDSLPKGVHYEKKKGHVLSYEGGMSVTQCSKEDRSSSG 1299
      478 -----ERGSAGS--KGSAGESGRPG-----EPGM-----PGQRGLTG 507
      1300 PPHETAAKRTYDMMEGRVGRALISSASIEGLMGRALIPPEHSHPHLKEQHHTIGSTIGI 1359
      508 PPGK-----QGRDGR-----PGPAGAPGPGNSGPGASGQGRGLGLV--GL 547
      1360 PRSYVEAOEDYLREAEALILKRETPPPPPSRDLTEAVYKTQALGPLKLPKPAHEGLVATVX 1419
      548 PGQOQ-----KEERGEDGQGSFGAPGLTGEKKGKEFGVAP--PGPG----- 591
      1420 EAGRSIHETPREELRHTPELPLARPLKEGSIIGTGTPLKYDTGASTGSKKHVRSILGS 1479
      592 SAGRGNGPGQOQ-----AGSMGPPGP-----PGASGDAQAQD---NGP 628
      1480 PGRTPPVNPLVDMAADARALERACYE-----SLKSRPGTASSSGGSIARGAPVLPVE 1532
      629 PGESGPEBGP-----GARG--ERGAPGERGPGGLTGAQGRKIGIRAGNYGHTGPPOQGE 682
      1533 LGKPRQSLTYEDHGAAPAGHL--PRGSPVYTREREPTPLQESLSSSKASQDRKLTSTPRE 1591
      683 MGPPGNVGL-----QGP--GELGSPGPPGANGPPGPGSGSPDPDPAGABGR----- 728
      1592 IAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSLPRGILPDLAAAYVL 1651
      729 -----GPVGP-----MGPSG-----PSGMGEAGD--- 748
      1652 PHLAPNTTYPHLYPPIYLRIGYPTALLENQTIINDYITSQQMHNHTATMAORADMLR 1711
      749 -----NEBPFGG-----APQGRD--- 763
      1712 GLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPTPTGTPTAMDRLAYLFTAPQ---P 1767
      764 ---PD-----LGPQ-----SPSPGAFGPPGRSGNPGQGBGLGP 796
      1768 FSSRHSSSPLSPGCTHLLTKPTTSSSRERDRDRERDRERERESILSTTYVHAPIW 1827
      797 TGARGETGPPGSGGTGDPGP-----OGBLG 822
      1828 RPTGSGSSGSSSGGCG-----GSSSRPASHAHQHSPLSPRTDALQORPSVLANTG 1882
      823 APQOQGERGETPQOQGPGRPTGSLGAPGAG--PGPTGPGSMAAGSPQPGARGBPG 879
      1883 MKGIITAVPSPKPTVLRST-----STSPVPRAATTPPATTCPLGCTLDGYVP 1930
      880 QSG-----SPQPGLAGRTGSGERGDKNQDSGPPPGPAPGAGSGSLGLAGG--- 931
      1931 TLMETVLLPKKAPRARPERPADGHAFLAKPPARSG-----LEPPASSP----- 1975
      932 -----SGPR--GPGGPAAGPPGAAGSRGAPGAGSGDRGSPGAVGAGNPGRAGENG 979
      1976 --SKGSEPPPLVPVSGHATITARTPAKNLAPHHASPDPPAP-----PASADPHREKTQS 2028
      980 PGSDGNDGAP--GPGSRGKEKDDTASGA--NGSPGAPGFIGALGAGAGSGPR----- 1026

```


QY 545 LKKEKTTDDTSGEDNDEKEAVASGKRTANSQGRKRKRITRSMANEANSEATITPOQSABL 604
DB 547 QAEBSSEKEG--SSEKEBGESEGETEA-----EABGEBA----- 580
QY 605 ASMEINSSGRWTEEMETAKGLLEHGRNWSAIAWVGSKTIVSOCKNFYFNKKQONLDE 664
DB 581 ---EAKKEKKVBEKESEVATK----- 600
QY 665 IIOQHKLKMEKERNARRKKKKAAPAAASEAAPPPVEDEEMASGVSGNEEMVEAEAL 724
DB 601 LVAD--AKVEKEBKAKSPVPKSP-----VEEKGS----- 628
QY 725 HASNGVPRGEGSGPATVNNSSDTSIPSPHTEAAKDTQNGPKPATLIGADPPRPPT 784
DB 629 ---PVPK-----SPVEKKG-----SPVPSKPVEEKSGSPVPS 659
QY 785 PPRTSRARIEPTPASEATGAPTPPAPPSAPPPV---VPKEKEETAAAPVE--E 839
DB 660 PVEBKGSFVPSKSPVEBKAKSPVPSKPVBEAKSKAEVKGEGQKEBEKEVKAPKEKYE 719
QY 840 GEEQKPPAAEELAVDTGKAEPVKSCTEEA----- 870
DB 720 KKEEKPKDVE---KKKAEFPKEBAVAEVVITKSVKMLEKETKEEKPLOQEKKE 775
QY 871 -----EEGPAG---KDAEAEATAEG--ALKAEKKEGSGGATTTAKSSGAPOD 914
DB 776 KAGGEGSESEESDDKAGKSRKEDIAVNGEVEGKEVEQETEKESGR-----EE 825
QY 915 SDSATCSAD--EVDAAEGGDKN 935
DB 826 EKGVTNGDLSPADKKGKGDKS 848

RESULT 83
T32008
Hypothetical protein K10G6.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32008
R:Davidson, S.; Wohlmann, P.; Mullen, G.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of *C. elegans* cosmid K10G6.
A:Reference number: Z21111
A:Accession: T32008
A:Stacus: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1819 <DAV>
A:Cross-references: EMBL:AF016669; PIDN:ABE6098.1; GSPDB:GN00020; CESP:K10G6.3
A:Experimental source: strain Bristol NZ; clone K10G6
A:Genetics:
A:Gene: CESP:K10G6.3
A:Map position: 2
A:introns: 41/2; 191/2; 556/2; 701/2; 1645/2; 1700/2; 1768/1; 1797/3

Query Match 2.5%; Score 329.5; DB 2; Length 1819;
Best Local Similarity 19.2%; Pred. No. 0.00022;
Matches 366; Conservative 228; Mismatches 679; Indels 637; Gaps 87;
QY 815 PSAPPVVVPEKEKEETAAAPVEEGE--EQKPPAAEELAVDTGKAEPVKSCTEEAEBG 873
DB 114 PQRPPPPQVDQSSSTSGSLPPLHYRYQLPPRPNNQRTPTQOLQYVK---VVEA--GG 169
QY 874 PAKGDAEAEAATVACALKAKEKSGGATTAASSGAP-----QDSDSATCSADEVD 927
DB 170 QAVAAQVQAQGS-----NRSGAAGVNSALQPKPLPLSLITSISSAAGSISAPS 221
QY 928 EAEAGDKRLLSPRSLTLPTGDP--ANAS-----PQKLDLQKQKQRA 971
DB 222 TSQPTSSLTSPSTSSSMAFKTTPNASSSSLIKQSQDVOEQQVDFEVANVSQ 281
QY 972 AIPPIQVTVHEPPPPREDAPTPKAPAPPPQNLQESDAPPOPGSSPR----- 1020

DB 282 IMSKNGKLVMEHREPLITGSLP--QLAELAPLP-----PKSGVYQCPNCNRLANARNLQRRH 336
QY 1021 ---GKSRGAPRADAEAPAEQKLPGRDPCMTSGLPFPVPRREVITKASPHAPDSAFY 1077
DB 337 QTCGSAQNAAP---QLAAMLQSRP--PCNSA-----PP---VAPTPASTSFQHN 379
QY 1078 APPGHPPLGLHDTA-----RPVLPRPPTISNPPPLISSA-----KHPSTLE 1119
DB 380 NSTGN-LTISYSSSSSRHOSLISYSPOLEHODLVGNPNWMLSGYEVKODPMLYQSPSGLS 438
QY 1120 ROIGALISQMSVOLHVPSSEAKAPVGVATGFLPLM--DP-KKLAPFSGKOEQLSPRQ 1177
DB 439 DSIM---SRDSSPHSPSPASHDQDMD--HLGFPDPLQPLHLHLSFSDADHKEKPREC 494
QY 1178 AGPPESLGV-PTAQAASVLR--GTALGSVPGSITKGIPISTFVPSDAITTYGSLTHGP 1234
DB 495 HEPDMLMTLDPTRPCGSERFYGINIDMP---LSDCDEPLMRSEASLSSSSQGRNP 551
QY 1235 ADVLYKGTITRLIGDSPRLDRGRDLSLPGHVIYEKKGHVLSEGGMSVTQCSKEDG 1294
DB 552 AAVFTRPPTTK----- 562
QY 1295 RSSGPPHETAAPKETDMMEGRVGSAISASIBGLMGRALPBRHSPHMLKEQHHR-- 1352
DB 563 ---RKPAKRPKSKKEASBEPKNSAILAAL--RKEPAAPQOPOLOFOQNYOPS 612
QY 1353 -----GSTQGIPIRSTVEAQEDYLKREAILKREGTPPPPPSRDLTEYKQALGP 1404
DB 613 POFQAPYGGSLPSTISASWLSASTSAAAA-----PERSEMTSDIVTSABNP 661
QY 1405 LALKPAHEGLVATVKEAGRSIHIEIPREELRHTPELP-----APRLKEGS 1450
DB 662 Y-----IHQHPQOPQOKSSPLEBLNEQESADDDDSRSSST 702
QY 1451 ITQGTPLKYDTGASTGSKNDVRSLLSGPGTFPPVHPLDVMAARALERAQYSESLKS 1510
DB 703 VENSITTT--TTATTTSSKS-----TGNP-----LFTGHCARQLCSMNLKR 743
QY 1511 RPGT---ASSSGGSIAR-----GAPVIVP-----ELGKPRQSPV---T 1542
DB 744 HRATKVAASSSNSNAASRPSPQSTPATPATPMLQASQADQPLQAPQOSMETTAVT 803
QY 1543 Y-EDGAPFAGHLPRGSPVTMEPTPRLQ---EGSLSSS-----KASQDKLITSTPREI 1592
DB 804 YTKTTPRESVANTWNTKRAQLISPKPRSGTITSESSSMTVVDALRAQHQKMQOQIOI 863
QY 1593 A-KSPHSTVPEHH-----PAPISPYEHLRGVSGVDLYRSHIPLAFPTSIPGCI 1641
DB 864 OFQOQOQOQRFQHHQOQOQAGRIIPRRPPPIINQVNPQOVQHNQHMPLR----- 918
QY 1642 PLDAAAAYTLPRHLAPNPPTYPPLRYLLRGYPTDALENROTIIINDYITSQMHNTAT 1701
DB 919 -----QPLQSPRP--PP-----PKGLLEHKT---DLVLI-----TSE 948
QY 1702 AMAORADMLRGSPRESSIALNYAAGPRGIIIDLSQVPLPVLPVPPGPTAMRLAYL 1761
DB 949 PLAEKMD-----AKRRS-----EGLVAVTSTPLPILQPRQSAQPARSROQOQOP 994
QY 1762 PTAPQ-PPSSRRHSSPLSPGPTLTKPTTT----- 1791
DB 995 PVAIVQVQFNGR---PLPPMQPLQPNPHNQOQHMLHQSOMNYQOVQOVQOVHQQO 1050
QY 1792 -----SSREKERDRER-----DRBREKSLITSTTT-----VE-H 1823
DB 1051 QMLQNHQOQHQQOHNQOQAPGNRSRSHSNVKNKEOEORQSGPLDSITISVPLSTIEVH 1110
QY 1824 APIWRPTGSSGSSGSSGSSGSSRPASHAHQHSPISRPTQDALQORPSV----- 1877
DB 1111 HHIMKRPGLQGOSSVDOSGTABEPKRAAS---QAVICECKTASRAKNAVGRHMA 1165
QY 1878 LHNTGKGIIT-----AVEP-----SKPTVLSTSTSSVPRPAATPTPTHCPGLG- 1923
DB 1166 VHKLTLDLIANPEQPLDPLSAVQAGRBHTVAGLETFDSALTKAPATKRSASEAPSAQV 1225

```

QY 952 RANASPOKEDLUKOLKORAAAIPIQVTKHNEPREDEAPTKAPAPPPQNIQPSDA 1011
DB 491 STTESSAPVTSTTESSAPV-----TPSSSTTESSA-PAPTPSSSTTESSA 540
QY 1012 P-----QQPSRPGKSRSPAPPADKEAPAAEQKLGDPCCWTCGLPFPVP 1058
DB 541 PVTSTTESSAPVPTPSSSTTESSSTVTSSTTESSAPV-PTPSSSTTESSAPVPTP 599
QY 1059 PREVIK-ASPHADPSAFVAPPGHPLPLGLHDTAPVLP-PTISNPPPLISSA 1112
DB 600 SSSSTTESSAPAPPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSS 658
QY 1113 KHPSTLEHQIGALISGCMVQLHVPVSEHAKAPVGVMTMGLPLPMDPKLAPPSGVKQQL 1172
DB 659 -----STTESSAPVPTPSSSTTESSAPVTS-----TTTESSAPVTSSTTSS 703
QY 1173 S-----PRQAGPPELSLGPVTAQEAASVLRGALSGVPGSITKGIPTRVPSDAITRG 1227
DB 704 SAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS 760
QY 1228 SITGTPADVLYKGTITRIIGEDSPSLDRGREDSPKHVYIGSKGHVLSYEGCMGV 1287
DB 761 SAPVPTPSS-----STT 772
QY 1288 QCKSED-GRSSSGPRHETAPKRTYDMMEGRVRAISSASIEGLMGRALPPRRHSPHLK 1346
DB 773 ESSSAPVPTPSSSTTESSAPVPT-----PSSSTTESSAPVPTPSSS----- 816
QY 1347 EQHNRGSIITOGIPRS--YVEAQEDYLRRKALKREKSTPPPPSRDLTEAYKQALGP 1404
DB 817 -----NITSAPSTPSSSTE-----SSSVPTPSSSTTES-----SSAP 853
QY 1405 LKLKRAHEGLVATYKEAGRSIHIEPREBLRHTPELPLAPRIKESITOGPRLKYDTGAS 1464
DB 854 VS-SESTESSAPV-----PTPSSSNITSSAPSIIP-SS 887
QY 1465 TTGSKGHVRLIGSPGRTPFPVHLDVMAADARALERACYEBSLKSRRGTASSGGSIAR 1524
DB 888 TTES-----FSTGTTVP-----SSSKYRGSQETISVS-ST 917
QY 1525 GAPVIVP-----ELCKPROSLPYE--DHGAPFAGHLPRG--SPVIMREPTPLQEGSLS 1576
DB 918 TETITVPKTTSTVTPSTTTITTTVCSTGINSAGETTSKCPKVTVTPTTTTSTVT 977
QY 1577 SKASODKLTSTPREIAKSPHSTVPEHHPHISPEYHLRGVGVLYRSHIPLAFDPTS 1636
DB 978 SSTT--TITTT--VCSTGINSAGE-----TTSKC-----SFKT 1006
QY 1637 IPRGIELDAAAYVILPHLAPRPTVPHLYPLINGVDTALENKQTHINDYITSQMH 1696
DB 1007 ITTVPCSTSPETASSTTSTSTTP-----VTTVSTTVVTTSTSTKPG 1053
QY 1697 HNTATAMQADMLRGLSPRESSLALNYAAGRGIILDSOVPHLVLVLPFGPATAMD 1756
DB 1054 GEITTFPTK-----NIPTY-----LTTIAPTSVTTVNTFTTTT 1092
QY 1757 RLAVYLPAPQPPSSHHSSPLSPGCPHUKPPTTSSSERERDRDRDREREKSLT 1816
DB 1093 TVCSGTGN--SAGETTSKCPKVT--TTPVCSGTGEVYTE-----ATTIVT 1136
QY 1817 S--TTTVEHAPIWRGTOSSG--SSGSSGGGSSSPASH-SHAHQSPISPTQAL 1871
DB 1137 TAVVTV-----VTTSTSTGNSAGKTTGYTTSKSVTTVTTLADAPVTPATNAV- 1188
QY 1872 QORPEVLN-----TGKGIITAVEPSKPTVLRSTSTSSPVRPATPEPATHCPLGTL 1925
DB 1189 ---PTTITTECSATNNAGETTSVCSAK-TIVSASAGENAPATTPVTTAIP----- 1239
QY 1926 DGVVPTLMEPVLLPREAPRVAPRPRADTGHAFIAPKPAR-SGLEPSSPEKSGEP-RP 1983
DB 1240 -----TVITTES--SVGTNSAGETTYGTSIPTTYITLIPGNGAKNYEVAT 1289
QY 1984 LVPPVSGHAT--IART-PAKONLAPHASDPAPAPASAS 2019

```

```

DB 1290 ATNPISIKTISQALATTASASVAPVWTSPLTGLQSHS 1328
RESULT 82
A27864
neurofilament triplet M protein - human
N|alternat|names: NF-M (medium) protein
C|species: Homo sapiens (man)
C|date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text change 13-Aug-1999
C|accession: A27864; A30157
R|Meyers, M.W.; Lazzarini, R.A.; Lee, V.M.Y.; Schlaepfer, W.W.; Nelson, D.L.
EMBO J. 6, 1617-1626, 1987
A|title: The human mid-size neurofilament subunit: a repeated protein sequence and the r
A|reference number: A27864; MUID:87275853; PMID:3608985
A|accession: A27864
A|molecule type: DNA
A|residues: 1-916 <MYE>
A|cross-references: GB:Y0067; NID:935045; PIDN:CA68276.1; PID:935046
R|Lee, V.M.Y.; Otvos Jr., L.; Garden, M.J.; Hollosi, M.; Dietzschold, B.; Lazzarini, R.A
Proc. Natl. Acad. Sci. U.S.A. 85, 1998-2002, 1988
A|title: Identification of the major multiphosphorylation site in mammalian neurofilament
A|reference number: A30157; MUID:88158120; PMID:2450354
A|contents: annotation; phosphorylation sites
C|genetics:
A|introns: 360/3; 402/2
C|superfamily: cytoskeletal keratin
C|keywords: coiled coil; phosphoprotein
F|615,628,641,654,680/Binding site: phosphate (Ser) (covalent) #status predicted
Query March 2.5%; Score 329.5; DB 2; Length 916;
Best local similarity 19.2%; Pred. No. 0.0001;
Matches 200; Conservative 149; Mismatches 361; Indels 333; Gaps 40;
QY 2 SGTOLVAQWRATEPRYPHSLSPVQIARTHTDVGILEYOHRSRDYASHLS----- 54
DB 30 SPSSGFRSQSMRSRSPSTVSSYKRSMLAPRLATVSSAMLSAESLDPGSSSLNGSG 89
QY 55 PGIITIQRRRPSLSLSEFQPNERSQELHLRPESHSLVPELGKSEMEFESKRRLLELP 114
DB 90 PGCGYKLSR-----SNEKEQLQGLNDRFAGYI--EKVHLEQNKKEIR-- 130
QY 115 DPLRPSPLATGPAGSEDLTKD-----RSLGKLEPVSPSPHPDPELELVPRLS 168
DB 131 -----HEIQLRKQKQSHAOGLDAYDOETBELATLBMVN-----HEKAQVQDSHIL- 178
QY 169 KEELIOWNDVDRREITWEOQISKLKKQQLBEEBAKPRPEKVPSP-IESKH 223
DB 179 EEDHLRKEEEREBARLDPTDEAIALRKQIEEASLVKVELDKVQSLQDEVAFLRSH 238
QY 224 RS-----LVQIITYENRKKAEMARILBGLCPQVEPLVNOPSDTROHENIKINQM- 276
DB 239 EEEVADLAIQ-----QASHIVE-----RKQYLKTDISTALKE 272
QY 277 -RKKLIVFKRNHARKOMKQFCORYDQLEALEKVERIENPRRRAKSKYREYEK 335
DB 273 IRSQLESISDNMOAEEM--FKCRYAKLLEAEONKEAI--RSAKI-ELAEYRKQ 323
QY 336 -----QPEIRKQRELOERQMSRVGQSGLSMSAARSEHVESEIIDGSEQEN----- 384
DB 324 LOSKSIELESRGKESLERQLSIE-----RHNDLSYQDTIQLENEELRG 373
QY 385 ---LEKQMR-----QNAVYPPMLYADDOQRIEFIMNGMLADPMKYVDKQ 427
DB 374 KWEAARHIREQDILLNVMALDIEIAAYRKL--EGEETFTSPAGSITGLPVTTHRPPI 430
QY 428 VMMNMSBEKEXTPFKPMQHPKNGFLASFLEKTVAE--CVLYUYTLKKENYKSLVR 485
DB 431 TISSKIQTKEAPILKVOHK---FYEBITIEETKVEDKSEMEBALTAITELAASKKE 486
QY 486 SYRRRGKQQQQQQQQQQQQQQQQQQPMPRSSQEKDEK-EKEKAKEKEKREVENDEK 544
DB 487 EKEKAERKEEPEABEEVEAKKSPVATATPYKEEGRGEBEGQEEBEEDEGAKSD 546

```


QY 804 GATP-----PRAPPSBAPPPV/PKEKEBETAAPVERGEBOKPPAAEELAVDQX 857
 Db 216 GARGQGGNGGEGEPGVSGPMGR-----GPP-----GPPKGP-----DDEE 256
 QY 858 AEPVSECTEBAEGPAKDAEAAEATAGALNAEKEGGSGRATYAKSGADPODS 917
 Db 257 AGKPKGA-----GERGPPOGARGPFG-PGLPVKGRGYPGLDGAKEGAP----- 305
 QY 918 SATCADEVEAEGDKNLLSPRSLLTPTGDPANASPOKPLDKOLKORAAIPIQ 977
 Db 306 -----GVKGE-----SSPGENGSP-GPMGRGLPGERGRTGAP 339
 QY 978 VTKHEPRREDAPTKAPAPPPONLOPESDAQOQSS-----PRGSR 1024
 Db 340 AAGAR---GNDQPPGAPGPPGVGPPGPPGAPAKGEAPGTARGEGAGQGGEG 396
 QY 1025 SPAPPADKEAFAAEAOCLPDDPCWTSGLFPVPREVIKASPHAPDPSAFSAPGHL 1084
 Db 397 TFGSGF-----PAGASGNG-----TDGTP-----GAKSAGAPGIA-----GAPFG 435
 QY 1085 PLGLHDTAPVLPREPTISNPPLISSAKHPSVLERQIGALISQGMSVOLHVPYSEHAKAP 1144
 Db 436 PFG-----PPD-----PQG-----ATGP 448
 QY 1145 VGPV-TMGLPLPMDPKKLAPFSGVQOEOUSPRGOAGRPESLGVPTPAQEAVALRTALGSV 1203
 Db 449 LQPKQOTGR-----GIAGFKGEQ-GPKGEPPGAPGAPGAPGAGEGKRG-ARGE 496
 QY 1204 PGGSITKGIPISTR-VPSDAITYRGSI-T-HGPADVLTKGTTIRLIGDSRSRL----- 1255
 Db 497 PGGVPIGPPEGRGAPGNGRPGODLAPKAP-----GERGSGLAGAPKG 543
 QY 1256 ---DGR-EDSLPKGHVIEGKKGHLVSEGMSVTQSKEDGSSSPRHETAA----- 1306
 Db 544 ANGDCORPGEPLRGARGL-TGAPGDA-GPOGKVGSGAPGABDGR--PBPPOGARGOP 599
 QY 1307 -----PKRTDMEGRVGR-----ISSASIEGLMR-----AIPERISPHILKQ 1348
 Db 600 GWMGPPGPKGA-----NGEPKAGEKGLPGAPGLRGLPGKGETGAPGPPGAPGERGE 655
 QY 1349 HHIRG-SITOGIPRSVYEAQEDYLREALKLKREGEPPP-----PPSRLT-EAVYTOA 1401
 Db 656 QGAPPSGFGLP-----GPPGPPGSGKRGDQGVGEAGAPBL 694
 QY 1402 LGPL-----KLKPAHEGLVATYKAGRSIHEIPREELRTPELPLAPRLKEGSITQ 1453
 Db 695 VGRBGRGPPGERGSPGAOQL-----QGRGLPRTGTGDPKASGAPGAPGAPGPGDQ 749
 QY 1454 GTPPKYDTGASITGSKHD-----VASLIGSPR-----TPPVHPLDVMADARALEBA 1502
 Db 750 GMP--GERGAGIAGAPKGRGDVGEGPPGAPKDGGRGLTGIGPPGAGANG----- 801
 QY 1503 CYESLKSPPGTASSSGSGSIARGAPVIVELGKPPROSLTYEDHGAPFAGHLPRGSPVTM 1562
 Db 802 --EKGEVGPAGSANG--ARGAPBERSETGPPGSGIA-----GPPGAD-----QPPGAK 848
 QY 1563 REPPTLQEGSLSSSKASQDRKLTSTREIANSPIHSTVEHHHPISPYEHLKGVGYD 1622
 Db 849 GEGGEGAGQKD-----AGAPGQGSAGAPGPGP-----TGVTGPK 884
 QY 1623 LTRSHPLAFDPSTISIRGILPLDAAAAYLPRHLAPRTYPLHLPPLILGYPDTALENR 1682
 Db 885 GARG--AOGP--PGATGPPGAPGAPGPPGSGNGPPG--PP--GPPGSGKDEP 928
 QY 1683 OTINDYITSQOMHNTATAMQORAMTLKGLSPRESSALINVAAG----- 1728
 Db 929 K-----GARGP--SGPPRAEBEPLOGAPGPPGKGERGDDGPPG 966
 QY 1729 -----RGIIIDL-----SOVPHL--PVLVPTPTGPTATMDRLAVLPTAPOP 1767
 Db 967 AEGPPGPGIAGRGIVGLPGGGRGPPGLPSPGSEPPGQAGPAGSAGR----- 1016

QY 1768 FSSRRSSPLSPGGETHLTKPRTTSSSERERDRDREREREKSLTSTTVENHAPIW 1827
 Db 1017 -----GPPGVPGRGLT-----GPPAG 1032
 QY 1828 PPGTEQSSSGSSSGGGSSSRPASHAHQHSFISPTODALOQPSVLTANTMKIT 1887
 Db 1033 EPGREGSPGADGPPGRDGA-----GVKG-- 1056
 QY 1888 TAVESKTVLRSTSTSPVPAATFPATHTCPIG-----GTLDDGVPLMEPV 1936
 Db 1057 -----DRGETGVGAPGAPGPPGSPGAPGTPGQDGRGAGAGPVPB----- 1100
 QY 1937 LLPKAPVAPABERERADTGH-----FLAKPPASGLEPASPSPKG 1978
 Db 1101 -GPARGIIGPPGGRGKGEGERGKLGKRGFTLOGLPBPGRPSGGQGSAPRPG 1159
 QY 1979 SEPRVLPVSGHATIAITPAKNTLAPHNASDPAPAPASADPHREKTQSKPFSIQEEL 2038
 Db 1160 SGPR-----GPPGPVPS----- 1172
 QY 2039 RSLGTHGSSYSPEGEVPVSPSLTTHDKGLPKHLELDKSHLEGELRPKQGPVXLGG 2098
 Db 1173 ---GKDGANGIIPGPIGPGRGRSGETGPAPGPN-----PBPGRPPGPG 1215
 QY 2099 EAAHLPHLRPLESOPSSSPLL-----QTARGVKHQHV-----TLAQHISEVITODYT 2148
 Db 1216 PGIDMSAFAGLGPGRKGDPLQWYADDAAGLRHDHAEVATLKSINQIESISPPGS 1275
 QY 2149 RHHPQOLSAPLAPLYSPGASCPVLDLRP--PSDLYLPPPDHAPARGSP--HSRGG 2203
 Db 1276 RKNPAR-----TCRDLKLCHPEWKSQGYWIDPNOCTLDAMKVCNMBTG 1320
 QY 2204 KSPREPNTSV-----LGS-----GEDEIEVSPREGTEBCH 2236
 Db 1321 ETCVYPNAPVAKKMMWSSKSEKHIWFGETINGFHFSTGDDNLAVYNAVQNT----- 1376
 QY 2237 SRSAVPLLYRDEQTEPSRMGSKSPGNTSOPAPFSLKLTESNASMYVSKKQOIKKLTNT 2296
 Db 1377 ---FLRLSTEGSN-----ITHCKNSIAYLDBAAGNLKALLI 1413
 QY 2297 HNRNPEYNISQPTGTFINMPAIT-----GTGLMTYRSQ 2330
 Db 1414 QSNDBVE--IRAEGNSRFTYTLAKDQCTKTKGKMGKTVIEYRSQ 1455
 RESULT 80
 CGHUIV
 collagen alpha 1(V) chain precursor - human
 N.Alternate names: procollagen alpha 1(V) chain
 C.Species: Homo sapiens (man)
 C.Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 16-Jun-2000
 C.Accession: S18802; S16024; A61142; S11303; S03378; S43642; S58665
 R.Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
 J. Biol. Chem. 266, 24727-24733, 1991
 A.Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e
 A.Reference number: S18802; MUID:92105142; PMID:1722213
 A.Accession: S18802
 A.Molecule type: mRNA
 A.Residues: 1-1838 <GRE>
 A.Cross-references: GB:M76729; NID:G189519; PID:AAA59933.1; PID:G189520
 J.Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaol, Y.; Kato, I.
 J. Biol. Chem. 266, 13124-13129, 1991
 A.Title: Complete primary structure of human collagen alpha-1(V) chain.
 A.Reference number: S16024; MUID:91302336; PMID:2071595
 A.Accession: S16024
 A.Molecule type: mRNA
 A.Residues: 1-81, 'OU', 84-389, 'A', 391-676, 'K', 678-1294, 'PS', 1297, 'RS', 1300-1553, 'R', 1555-
 A.Cross-references: GB:D90279; NID:G219509; PID:BA14323.1; PID:G219510
 A.Note: parts of this sequence were determined by protein sequencing
 R.Yaol, Y.; Hashimoto, K.; Takahara, K.; Kato, I.
 Exp. Cell Res. 194, 180-185, 1991
 A.Title: Insulin binds to type V collagen with retention of mitogenic activity.
 A.Reference number: A61142; MUID:91224163; PMID:1709100

Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
 A>Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pro
 A:Reference number: A94227; MUID:90370826; PMID:1975693
 A:Accession: A33116
 A:Molecule type: DNA
 A:Residues: 171-172, 'C', 174-175 <ATA>
 A>Note: mutant sequence from a family with family with primary generalized osteoarthritis
 R:Diab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996
 A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
 A:Reference number: S64673; MUID:96195147; PMID:860302
 A:Accession: S64674
 A:Molecule type: Protein
 R:Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
 Eur. J. Biochem. 234, 125-131, 1995
 A>Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cat
 A:Reference number: S63514; MUID:96096730; PMID:8529631
 A:Accession: S63514
 A:Molecule type: Protein
 A:Residues: 243-261, 575-590, 756-763, 'X', 765-779 <FRA>
 R:Tiller, G.E.; Weis, M.A.; Colombo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
 Am. J. Hum. Genet. 56, 388-395, 1995
 A>Title: An RNA-splicing mutation (G+51VS20) in the type II collagen gene (COL2A1) in a
 A:Reference number: 138867; MUID:95150028; PMID:7847372
 A:Accession: 138867
 A:Structure: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TIL1>
 A:Cross-references: EMBL:X15195; NID:9557053; PIDN:AAB60370.1; PID:9557054
 R:Ramirez, F.
 submitted to the EMBL Data Library, December 1988
 A:Reference number: S04892
 A:Accession: S04892
 A:Molecule type: mRNA
 A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
 A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
 R:Viikula, M.; Peltonen, L.
 FEBS Lett. 250, 171-174, 1989
 A>Title: Structural analyses of the polymorphic area in type II collagen gene.
 A:Reference number: S05000; MUID:89325561; PMID:2753125
 A:Accession: S05000
 A:Molecule type: DNA
 A:Residues: 630-640, 'A', 642-785 <VIK2>
 A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
 PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
 R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
 J. Biol. Chem. 267, 22522-22526, 1992
 A>Title: An amino acid substitution (Gly853->Glu) in the collagen alpha 1(I) chain pro
 A:Reference number: A44309; MUID:93054548; PMID:1423602
 A:Accession: A44309
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA; mRNA
 A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
 A:Cross-references: GB:L00977; NID:9180812; PIDN:AAB3914.1; PID:9258774
 A>Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence wer
 A>Note: this translation is not annotated and this publication is not cited in GenBank
 R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
 A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
 A:Reference number: S16502; MUID:90251662; PMID:2339128
 A:Accession: S16502
 A:Molecule type: DNA
 A:Residues: 1164-1184, 'GPGKDGANGIPGP', 1185-1199 <TIL2>
 A:Cross-references: EMBL:M37126; NID:9180808; PIDN:AAA52037.1; PID:g180809
 R:Chenah, K.S.B.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 2555-2559, 1985
 A>Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A:Reference number: A02858; MUID:85190534; PMID:3857598
 A:Accession: A02858
 A:Molecule type: DNA

A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>
 A:Cross-references: GB:J00116; NID:g180395; PIDN:AAA5197.1; PID:g180396
 R:Eljima, K.; Viorio, T.; Viorio, E.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
 A:Reference number: A27280; MUID:88067771; PMID:2825137
 A:Accession: A27280
 A:Molecule type: DNA; mRNA
 A:Residues: 1175-1487 <ELI>
 A:Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:g30097
 A:Experimental source: fetal epiphyseal cartilage
 R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: Protein
 A:Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408 <VAN>
 A>Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
 R:Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A>Title: Isolation and characterization of genomic clones corresponding to the human typ
 A:Reference number: A21733; MUID:8418798; PMID:6320112
 A:Accession: A21733
 A:Molecule type: DNA
 A:Residues: 1245-1295 <STR1>
 A:Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:g4378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909, 'PE', <STR2>
 A:Cross-references: GB:X01785; NID:930035; PIDN:CAA25082.1; PID:g1335032
 R:Nunez, A.M.; Francopano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A>Title: Isolation and partial characterization of genomic clones coding for a human pro
 gene.
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358 <NUN2>
 A:Cross-references: GB:M12048; NID:g180017
 A>Note: this translation is not annotated in GenBank entry HMWCT2A, release 111.0
 A>Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
 R:Sangiorji, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
 Nucleic Acids Res. 13, 2207-2225, 1985
 A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
 A:Reference number: 137249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7-28, 'R', 99-114, 541-578, 786-802, 1055-1056, 'N', 1058-1068, 'T', 1070-1109, 1200-1
 A:Accession: 184453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 7-28 <SAN2>
 A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
 A>Note: the GenBank PID is based on an incorrect reading frame
 A:Accession: 137250
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 541-560 <SAN2>
 A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
 A:Accession: 137251

Query Match 2.5%; Score 333.5; DB 1; Length 1487;
 Best Local Similarity 20.4%; Pred. No. 0.00013;
 Matches 377; Conservative 116; Mismatches 628; Indels 723; Gaps 85;

Qy	726	ASGNEVPGECSGATVNNSSDTSIPSPHTEAK-----DTGQNG-EPK-----P	770
Db	96	ASGPGPGGQGGEGEDIDIVGPKGPPGQSPAGEGGRGGRGKGAGAPGRGDSFP	155
Qy	771	ATLGADGP--DPGPTPP-----RTSRAPTEPTPASEAT	803
Db	156	GTLAGNPGPPGPPGPPGGLGCGNPAQWAGGFDKAGAGQLGTVWGPGWPGGPPGPPGA	215


```

Db 408 -----EEDGVRCGAPLARAGPPSP-----APAAP---RP 437
Qy 1410 AHEGLVATYKEAGRSIHETPREELRNTPELPLAPRLKSGSTTQCTPLKYDPGASTGSK 1469
Db 438 SASASANTSSSAASAPAPAP-----PARPPKRSRTNNHLSLMADPPPTDG--- 486
Qy 1470 KHDVSLISPORTPPVPLDVMDARALERACYESISKSRPG-----TASSGGS 1521
Db 487 -----PLTLPGEPPWGSPP---PADGRVR-----YGAADDSREGLMDEDVDVQAARVYA 534
Qy 1522 IARGAPVIVBELCKPQOS-----PLTYEDHGAFFAGHLPRGSPVTVREPTPLIQEGSLSS 1576
Db 535 AAGPVAVFPIPEWGDSSKQHEALVRLIYS--GA-----AGEAMSWLQ--NPRMQ----- 578
Qy 1577 SAASODRKLTSTPRELAKSPHS-----TYPEHH-----PHPISP--- 1610
Db 579 ---APQRFNPOFCQRRVVAHGHGSGFITSGVTPPEPLHIGDMAAODPLMALPPHVASAVAM 635
Qy 1611 ---YEHLRGVSGVDLYRSHIPLAF-----DP-----TSIPRGPDL 1643
Db 636 SRRYDPTQXTFLLQSLRRAVADMAVPGRAADPRAGATVEALCARVAAFAAAQPGRVVR 695
Qy 1644 DAAAAAYTL-----PRHLAPNPTYPHLYPEYLLRGYPDTAALENQTIINDYIT 1691
Db 696 ELADACVLACRGVLERLLPCPLRLPAPARAPALGPACLEEV--TALL-----LALRDALP 749
Qy 1692 S-----QOMHNTATMAQORADMLR---GLSPRESSLALNYAA--GPRGID--LSQ 1736
Db 750 GAGPAPERQOADSVALVARTVAFLVYVSDGARARA--AMTYAALFAPANVAGARLAE 807
Qy 1737 VPHLPVLVPTPTGTPATMDRLAFLPTAPQPFSSRRSSPLSPGCGTHLTKPTTSSSR 1796
Db 808 AARAPAPAPAGLPPLWPEQEGVLVPAAP--AAAGAPGGLRSGG---SSPASTSG-- 861
Qy 1797 EEDRRERDREREREKSLTSTTVENHAPIMRPT-----EQSSG 1836
Db 862 -----SSTKSSSGTSGLSGSGVYARLPRRRRGPASARAQEBAPARAGARRPDGDEHDG 915
Qy 1837 SGGSSGGGGSSSRPASHAHOHSPISRTODALOQRSVLNHTMKGIITAVESKPT 1896
Db 916 LGGSLARGDG-----HGRDDEEDRGR--RKRSLL-----GLGPAPDAPAL 956
Qy 1897 VLKSTSS-----PVPAATFP-----PA--THC--PLGGLTDGVYPTLMRP 1935
Db 957 VSSSSSSSEDDRLRLRPLGMPENHAPDGGFRRVAGETHPRPAAALAAVCP----- 1011
Qy 1936 VLLPKAPVVAR-----PR--RPADDTGNAFLAKPPASGLEBPASSPKSGSEPRPLV 1985
Db 1012 -----PEVARALVDQEVFPELWRPALTFDPALLAHIAARGAAGAPLRRRAAMWRQIA 1064
Qy 1986 PPVVS-----GHATTARTPAKTLAHHASPPDPAPPA--SASDPHEKT----- 2026
Db 1065 DEDVAVRVVLYDPLPHEELCAEPAEAG--APRGGLPAPRGSLALLAAPHNLCTPDISHAMAG 1123
Qy 2027 -----QSKPFSIQELESLSLGYHSSYSPEGEVPS-----P 2058
Db 1124 NWTGPRDITRLAAQVLLISARDLSAGA-----VEYICSLRGARRRLIULDTIEDMP 1177
Qy 2059 VSSPSLTMDKGLPKHLELDKSHLBELRPPKQPGVYKQGEAAHLPHLRPLPESQSSP 2118
Db 1178 ADGPAAVG-----DYHV-----YVRARLDPAQCAVR-----WPECRELAVALVDSSS 1219
Qy 2119 LLIQTA-----FGVKGHQVVTLAQHISVITODYTRHHPOOLS--APLPP 2162
Db 1220 IYGPACFARVEASPARLHPGAE-----LTLCSQDNNVRYVSTRAGPRTVPPLP 1268
Qy 2163 LVSFPACGPVL---DLRRPSDLYLPPPDHGAAPRAGSPHSGCKRSPKSPKNTSVLGGG 2218
Db 1269 PRAIYRRTVPTVDCGCDMARORSALDGRPDPGAGAFGRANRWGLGAPLRPVVVS CG 1328
Qy 2219 EDGIEIVSPPEQTEGHSRSAYVLLYEDGEGTEPSRSGSKSPGNTSOPPA 2270
Db 1329 RGLAELRREPGI--PAELRAFCAALLLBPDAEAAVLVLAPGALLAAAGAPPA 1378

```

RESULT 79

CGHUC6

collagen alpha 1(II) chain precursor [validated] - human

N/alternate names: procollagen alpha 1(II) chain

M/contains: chondrocalcin, collagen alpha 1(II) chain precursor splice form 1; collagen

C/Species: Homo sapiens (man)

C/date: 26-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000

C/accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63

7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910

R/Ryan, M.C.; Sieraski, M.; Sandell, L.J.

Genomics 8, 41-48, 1990

A/Title: The human type II procollagen gene: identification of an additional protein-cod

A/Reference number: A38513; PMID:91184811; PMID:2081599

A/Accession: A38513

A/Molecule type: DNA

A/Residues: 1-103 <RYA>

A/Cross-references: GB:M60299; NID:G180883; PIDN:AAA73873.1; PID:G180884

R/Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.

Nucleic Acids Res. 17, 9473, 1989

A/Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla

A/Reference number: S06715; PMID:90067946; PMID:2587267

A/Accession: S06715

A/Molecule type: mRNA

A/Residues: 1-28, 'R', 99-1487 <SU2>

A/Cross-references: EMBL:X16468; NID:G29515; PIDN:CAA34488.1; PID:G29516

A/Note: alternative splice form 1

R/Viikula, J.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.

Biochem. J. 285, 287-294, 1992

A/Title: Structural analysis of the regulatory elements of the type-II procollagen gene.

A/Reference number: S24270; PMID:92344585; PMID:1637314

A/Accession: S24270

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-28 <VIK>

A/Cross-references: EMBL:X58709; GB:S40537; NID:G35659

A/Note: this translation is not annotated in Genbank entry HSPROCOE1, release 111.0

R/hunz, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.

Gene 44, 11-16, 1986

A/Title: Promoter region of the human pro-alpha-1(II)-collagen gene.

A/Reference number: A24828; PMID:87031574; PMID:3021582

A/Accession: A24828

A/Molecule type: DNA

A/Residues: 1-8, 'T', 10-28 <NUN>

A/Cross-references: GB:M25698; NID:G180872; PIDN:AAA52051.1; PID:G55237

R/Balwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.

Biochem. J. 262, 521-528, 1989

A/Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)

A/Reference number: S06496; PMID:90026318; PMID:2803268

A/Accession: S06496

A/Molecule type: mRNA

A/Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F'

A/Cross-references: EMBL:X16711; NID:G30040; PIDN:CAA34683.1; PID:G30041

A/Note: alternative splice form 1

R/Ryan, M.C.; Sandell, L.J.

J. Biol. Chem. 265, 10334-10339, 1990

A/Title: Differential expression of a cysteine-rich domain in the amino-terminal propept

A/Reference number: A35428; PMID:90285153; PMID:2355003

A/Accession: A35428

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 27-81, 'V', 83-103 <RYA2>

A/Note: alternative splice form 2; splicing appears to be under developmental regulation

R/Su, M.W.; Benson-Chanda, V.; Visseling, H.; Ramirez, F.

Genomics 4, 438-441, 1989

A/Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf

A/Reference number: A30147; PMID:89233138; PMID:2714801

A/Accession: A30147

A/Molecule type: DNA

A/Residues: 104-157, 'P', 159-236 <SUN>

A/Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB

R/Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.

Db 223 AGAP-----GVKGE-----SGSPGKNGSP-GRMGRLPG 261

Qy 969 RAAALPRLQVTKVHEPPREDAPARTKRAPAPPQNLQPSDAPQPPSS-----1018

Db 262 ERGRTGPGAGAGAR--GNDGQRPAGPFGPVGAGGPGFPGAPFAKABAGTGAARGEG 318

Qy 1019 ---PRGKSRSPAPADKEAFAAEOKLPDPPCMTWSSGLPFVPVPEVIKASPHADPSAF 1075

Db 319 ACPGPGEGTPESSPG-----PAGAAGNPG-----TDGIP-----GAKGASGAGIA--359

Qy 1076 SYAPGPHPLGLHDTAPVLPRLPTISNPRLISSAKHPSVLERQIGALISQMSVOLHV 1135

Db 360 --GAGFGGPRG-----369

Qy 1136 PYSHAKAPVGVNVMGLPMDPKLA---PESGVKOBOLSPRGAGPESLGVPTAOEA 1192

Db 370 -----PPGP--OGATGGLGPKQTGEPGIGAGRGD--GPKGEPGPGAGPAGPAGE 418

Qy 1193 SVLRGTALGSVVGSGSITKGIPSTR--VPSDAITVGSIT--HGTADVLTKGTITRIIGE 1249

Db 419 EGKRG--ANGP--PGAGPVGPRPGERGAPGNRGPRGQDGLAGPKAR-----GE 463

Qy 1250 DSPSL-----DRGR--EDSLPKGHVIEGKKGHVLSYEGGMSVTQCSKEDRSSSG 1299

Db 464 RGPGLAGPKGANGDPRGPRGEBGLPGARGL--TGRPGDA--GPGVKGPGSAGPAGEDGR--PG 519

Qy 1300 PHENLA-----PKRTYDMMEGVRA-----ISSASIGLMOR-----AIP 1337

Db 520 PPGPGAGGQPGWMPGPGPKA---NGPRGAGKGLPGAGLGLPKQDGETGAAGP 575

Qy 1338 ERHSPHLKEOHNRIG--SITGQIPRSYVBAQDYLRREAKLKEGTEPPR-----PESR 1391

Db 576 GPAGAGERGEAGRGPSGQGLP-----GPRPPEGGKPRDQ 614

Qy 1392 DLT-EAYKTALGPL-----KLKRAHEGLVATVKEAGRSIHETPRELHTEPLIA 1442

Db 615 GVPGEAGAPGLVGPGERGPRGERGSPGAQGL-----QCARGLPTGTPGCKKASGAG 669

Qy 1443 PRPLEGSLTGTPTPLKYDTGASTGSKND-----VRSLSGPR-----TFPVNPLD 1491

Db 670 PPGAGGPPGLQOMP--GERGAAGIAGPKDRDQVGEKGEPAKGDGKGGLGIPGPRP 727

Qy 1492 VVADRALERACYESLSKSRPGTASSGSGSIRGAPVIVPELGEKROSPLTVEDHGAPPA 1551

Db 728 AGANG-----EKGEVPRPPAGTAG--ARKAPERGEGETGP--GP-----AGPA 767

Qy 1552 GHLPRGSPVTMEPTPRLQEGSLSSSKASQDRKLTSTREIAKSPHSTVPEHHPRIPBY 1611

Db 768 G--PGQAD--GQPAKKEQG-----EAGQ-----KQDAGARQPGPSGAPGQGP-808

Qy 1612 EHLLAGVSVDLVRSIHPLAPDPSTIPRGILPDAAAYVLPHNLAPNPTPHLYPLILR 1671

Db 809 ---TGVTGPKKARG---AAGP--PGATGPGGAAGVGPGSNGNPG-----PP-----848

Qy 1672 GYPTDAALENROTIIINDYITSOOMHNNTATAMAQRADMRLGSPRESSLALUYAAGP---1728

Db 849 GPPGSGKDGPK-----GARGD--SGPRGAAGDGLGPRAGPBE 886

Qy 1729 -----RGIIDL-----SOVPHL--PVLVPTGTPATAMD 1756

Db 887 KGPBGDDGSGPDGPPPGQGLAGQRIYGLPQORGERGPRGLPGPSGEPGKQAGASGD 946

Qy 1757 RLALYLPATQPPSSSRHSSPLSPGCTPHLTKPTTSSSRERDRDRERDREREKSILIT 1816

Db 947 R-----GPPGVGPRGLGP---AGPRGREGTPGAAGPGRGAGAV 985

Qy 1817 STTVEHAPITMPGTBOSSGSSGSGGSSSR-----PASHENH--QNSPLSP 1865

Db 986 KQDRGEAGLAPGARGPSPGSPFQAPRTKQGDKGAGAQGSMGRAPGAGLGPQGP 1045

Qy 1866 RTQDA---LQORPSVLHN--TGMKGITAVEPSKPTVLSTSTSSPVRAA--TEPPATH 1918

```

Db      1046  RGDKGAGEAGEGRLKGRHGTGLQGL-----PCPPPSDDQAGSAGPAGSPGPPGPPV 1100
QY      1919  CPLGGTLDGVYPTLMEVLLPKKAPRVARDEPRADTGHAFLLAKPARPSGLPASPSSKG 1978
Db      1101  GPSSG-XDGA-----NGIRGPIGRPPRGGRGGTGGPAGPPGNPG----- 1137
QY      1979  SEPPRLVPVSGHATARTPAKNLAPHHASBDP-----PAPPASADPHREKTOSKEPST 2033
Db      1138  -PPGPPGPPGCG--IMSAFAGLGPKEKGGDPLQVRADEAAGLAPRHDEEV----- 1186
QY      2034  QELERLSLGYHGSYV-SPEG 2052
Db      1187  -EATLKSLNNQIESIRSPG 1205

RESULT 78
EDB31P
Immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)
C:Species: suid herpesvirus 1
C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997
C:Accession: S04713
R:Chang, A.K.
Nucleic Acids Res. 17, 4637-4646, 1989
Article: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies v
A:Reference number: S04713; MUID:89315207; PMID:2546124
A:Molecule type: DNA
A:Residues: 1-1460 <CHE>
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation

Query Match      2.5%; Score 335; DB 1; Length 1460;
Best Local Similarity 20.4%; Pred. No. 0.0011;
Matches 341; Conservative 140; Mismatches 643; Indels 548; Gaps 71;

QY      832  AARPVEGEFOKPPAAEELAVDTGKAEFPYKSECTBEAEGCPA-----KKQDAEAE-A 885
Db      22  AAAAAEEEGIGASGP-----DCGSGGSRRRGSGGEBLLFGPGGLPSDDAAEAEAATLA 73
QY      886  TAGGLAKKKKEGSGSGRATTAKSSG---APQDSSSATCSADEVDEAGCDKXNLLSPRP 942
Db      74  AAGGATRRPPPSAQQQKHARSGSEIVLDDDE---EEDEPGSPAGSPVGLSTRAP 129
QY      943  SLTPPTGPPRANASPOKRLDLKOLKORAAAIPIQ--VTKNHPPREDAAETKEPARAP 1000
Db      130  STVSSSSPGCGPARGPGRRRRQHSQKRRPPRAARGARRPPRRRRPPRRAPARAP 189
QY      1001  PPQNLQPESDAPQOFGS---SPRGKSRSPARADKEAFAAEOQLPGDPFCWTGGLPPPV 1057
Db      190  APR--RPRGDDPRPGGTGTSVSP--GRRRGLGRHHQHSQQWRPQRHNGGSP-----LPQR 241
QY      1058  PPRVETKASPHARPDSASVAPRGHRLPLGLHDPARVLPRLPPRLTISNPRLLISSAKHPSV 1117
Db      242  PP-----PGSRPPRAAAPPAAGTAVVTTTSTASFWLDEPARAARLIDPAAAMPREBL 295
QY      1118  LERQIGAIQGSMSVOLAHVYSEHAKAPVGTWGLPLPMDCKLAPFSGVQEOQLSPRGQ 1177
Db      296  LQPG-----LQDLNHHRRRRARP-----RPRG-----RGR 323
QY      1178  AGPPESLQVP-----TAQEAIVLRTALGSVPGGSITGCIPTSTVPSSAITYRGST 1229
Db      324  TRPRGRGAPLQQRPRRRRAGEGALLRGGRGSSSSSGG----- 361
QY      1230  THGTADVLYKGTITRIIGEDSPSRLDGRGDSLPKHGVITYEGKKHVLSEGGMSVTC 1289
Db      362  -----SDSLSPARSPAPR----- 376
QY      1290  SKEDRSSSGPHEHTAARTYTDMMEGRVGSAISASITGLMGRALIPRRHSPHLLKEQH 1349
Db      377  -----AAAAAAARRSASSSSSSSSSSSSSSEG----- 407
QY      1350  HIRSGITGIGIPRSYVEAOEDYLRBAKLKKEGTPPPPPPSRDLTEAVKTOALGPKLKP 1409

```

Matches 330; Conservative 117; Mismatches 613; Indels 474; Gaps 75;

```

QY 725 HASGNEVREGSCGATVNN--SDTE-----SIPSPHTEAKDTGONGPKRPA 771
Db 35 HAS--STGEGKETATQSSVSPSTEEKNAVSMTSVLSHSPGSGSTTGQODVTLAPA 92
QY 772 TLGADG-----PPPGPTPPRR--TSRAPIEPTPASEA-----TGAPT 807
Db 93 TEPAGSAATWGDVTSVPTRPALGSTTPPAHDVTSAPDNKPAAGSTAPPAHGVTSAPD 152
QY 808 PPPAPPSBAPPPVVPKEKEETAAAPVEEGEQKPAAMELVVDGKA---EPPKS 864
Db 153 TRPAGGSTAPPAHGV---TSAPDTRPA---GSTAPPAHGVTSAPDTRPAAGSTAPPAH 205
QY 865 ECTEEAEBSGAKGDAAEAATAGALKAEKKGSGRATTAKSSGAPODSDSATCSAD 924
Db 206 GVTSAAPDTRPAAGSTAPPA---HGVTSAADTRPAAGSTAPPAHGVTSAPD 252
QY 925 EVDABGDKNRLSPRSLTPTGDDPRANASPOKPLDKOLKQRAAALPPIQ-VTKYHE 983
Db 253 -----TRPAGSTAPPAHGVTSAPDTRPA---GSTAPPAHGV--- 288
QY 984 PRREDAAPTTPAP-PAPEPPQULQPESDAPQOGSGSPGKSRPAPPADEKAFAPAAQXL 1042
Db 289 ---SAPDTRPAAGSTAPPAHGVTSAPDTRPAAGST---APPAHGVTSAPDTRPA 336
QY 1043 PGDPPCMTSGLPFPVPPREVIRAKSPHAPDPSAFYAPGHPPLGLMDTAPVLPPTT 1102
Db 337 PGS-----TAPPAHGVTSAPDTR-PPAGSTAPPAHGV---TSAPDTRPAAGST 380
QY 1103 SNPPPLISAKHPSVLERQIGALISQMSVOLHVPSYSENAKAPVGTWMLPLPMDPKLA 1162
Db 381 APPPAHGVTSAP-----PDRPAGSTAPPAHGVTSAPDTRPA-A 416
QY 1163 PPSGVQEOQLSPRGAGPESIGVPTAOBASVLRGTALGVSFGSITTKGIPSTR-VPSDS 1221
Db 417 P-----GSTAP-APAGVTSAPDTRPAAGST-APPAHGVTSAP-PTTRPAAGST 460
QY 1222 AITRGSTHGTTPADVLYKGTITRIIGEDSPSLDRGRDLSLPKHVITYEGKKHVLSTY 1281
Db 461 A-----PPAAGVTSAPDTR-----PAFG-----STAPPAHGV----- 487
QY 1282 GGMSTVQCKEKDQSSSGPHEHAAKRTYDMMEGVGALASISLEGMAIPERHS 1341
Db 488 -----TSAPDTRPA---PGSTAP-----ANGVTSAP-----PTTRPA 516
QY 1342 PHHLKEQHHRIGSITGIPRSYVEAOEDYLREAKLLKEGTTPPP-----PSRDLTAY 1397
Db 517 P-----GSTAP-PAHGVTSAPD-----TRPAGSTAPPAHGVTSAP 551
QY 1398 KTQALGPLKLPRAHEGLVATVKEAGRSIHETPREELRHTPELPLARPLKEGSIQTGPTL 1457
Db 552 DTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPA--- 597
QY 1458 KYDTGASTGSKGDVRSILG---SPQRTPPVHPHPIVDVADALERACEYSEIKSRPST 1514
Db 598 -----GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTR-----PAPGS 639
QY 1515 ASSSGSISARGAVIYVELGKPRQSPITYEDHGAAPPAHLPGRSPVTMBEPTRLOEGSL 1574
Db 640 TAPPAHGV-TSAPDTRPAAGST-----APPA-HGVTSAPDTR-PPA- 677
QY 1575 SSSKASODKLTSTPREIAKSPHSTYVPHHPHPISEPHLLRVSQVDLYRSIIPLAFP 1634
Db 678 -GSTAPPAHGVTSAP-DTRPAAGSTAPPAHGVTSAPDTRPAAG-----STAPPAHGV 727
QY 1635 TSIPRGIPLDAAAYVLPRLHAPNPTVPHLYPYLLRGVPTDALENQTIINDYITSSQ 1694
Db 728 TSAPDTRP-----AGSTAP-PAHGVTSAPDTRPAAGSTAPPAHGVTSAP 771
QY 1695 MHNHTATAMAQRADMTLGLSPRESSLALNTAAGRGIIIDISQVPHLPVLPPTGTPATA 1754
Db 772 ---DTRPAAGSTAPPAHGVTSAPDTR-----RPAAGSTAP 803

```

```

QY 1755 MDRLATIP-TAPQFSSRHSSPSLSPGPTH--LTKEPTTSSSRERDRDRERERE 1811
Db 804 AHGVTSAPDTRPA---GSTAP-----PAHGVTSAPDTRPA--- 837
QY 1812 KSLITSTTVSHAPIMRPTGEOSSGSGSGGGGSS--SRPASHAHQHSPISTPTOD 1869
Db 838 -----GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTS 889
QY 1870 ALQORPSVLIHTGKGIITAVEPSKPTVLNSTSSSVRPAATPPATHTCEPLGGLDGVY 1929
Db 890 APDTRPAAGSTAPPAHGVTSAPDTRP-----APGSTAPPAHGVTSAPDTRPAAGSTAPPAH 945
QY 1930 PTLMEPVLPKEARVARPREPRADTGHAFIAXPAR-----SGLEPA-----SSPSKSEP 1981
Db 946 GVTSAAPDTRP-AGSTAP-----PAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPA 997
QY 1982 RPLVPVSGHATIAIT--PAKNLAPHN---ASPD-PPAPASADPHREXTQSKPSTIQ 2034
Db 998 GSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDNNRPAAGSTAPPAHVNVTASGASAGS 1057
QY 2035 ELEIRSLCYHSSYSPEGVEVVS--PVSSPSLTHDKGLPHHLELDKSHLEGELRPQP 2091
Db 1058 ASTL-----VHNGTSARATTTTPASKSTPSTPS--HSDPTTTL---ASH----- 1097
QY 2092 GPVILGGEAAHLPHLRPLPESQPSPLQOTAPGVKGHQRVTLAHISEVITQDYTRHH 2151
Db 1098 -STKIDASTHSTVPLTSSNHSSTPOLST--GVS-----FFLSFHSNL----- 1141
QY 2152 PQOLSAPLPAPLYSPFPGASCFLDLRRPSDLYL 2185
Db 1142 --QFNLSLEDPSDY-----YQELQORDISEMFL 1167

```

RESULT 77
T45467
collagen alpha 1(ii) chain precursor [imported] - horse
N.Alternate names: type II collagen
C.Species: Equus caballus (domestic horse)
C.Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C.Accession: T45467
R.Charidson, D.W.; Dodge, G.R.
submitted to the EMBL data library, June 1996
A.Description: Cloning of equine type II collagen and modulation of its expression in eq
A.Reference number: 222977
A.Accession: T45467
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1418 <RIC>
A.Cross-references: EMBL:U62528; PDB:AAB05773.1
C.Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology,

Query Match 2.5%; Score 335; DB 2; Length 1418;
Best Local Similarity 22.4%; Pred. No. 0.00011;
Matches 340; Conservative 93; Mismatches 567; Indels 520; Gaps 81;

```

QY 721 AEAHASGNEV-----PRGSCGATVNNSSDTEISIPSPHTEAK-----DTGQNG-P 767
Db 18 AAVLKHQGDVAVQPKGQKGRPDIXOIVGKPPRGQGRGEGGPRGDRDKGEKAP 77
QY 768 KP-----PATLGADGP-PPGPTTP-----RTSRAP 794
Db 78 GRRGRDGEPTGNGPGRPGPPGPGIGNFPAQMAAGFDEKAGAGQMGVQGPWGM 137
QY 795 EPTPASEATGATP-----PPAPSPAPPPVVPKEKEETAAAPVEEGEQKPPAA 848
Db 138 GRRGPPRGAGAGPQGFQGNPGEPPGEPVSGWGR-----GPP-----GPPKPG-- 183
QY 849 EELAVDTGAEBEPVSECTEEAEBSGPAKGDAAEAATAGALKAEKKGSGRATTAKS 908
Db 184 -----DDGEAGKRGKS-----GERGPPGQAGRGFPPT-PLGVYKGRHGVGLDGAKE 232
QY 909 SGAPODSDSATCSADEVDABEGDKNRLSPRSLTPTGDDPRANASPOKPLDKOLKQ 968

```

Db 2041 -----TITATGTTNPESTPTGTTPIPVLTITTAATTAATSTVTPS--SALGTH----- 2088

Qy 2086 LRPQPGCVKLGGEAHLPHLRPLESQPSSPLQTR--PGVKQHVVTLAQHISVYI 2143

Db 2089 -TPVP-----NTTATTHGRSLPPSSPHTVPTAMTSATGIIIG-----TTHITE-- 2131

Qy 2144 TQDTRHHHQQLSLAPLPLVSPGASCPVLDLRRPPEDVLPPDHGAPARGSPHSHGG 2203

Db 2132 -----PSTGTHPTAATGTTGTOPSTPALSSPHPSRTTBSPSPGTTTGG--HRTGT 2181

Qy 2204 KR-----SPENKTSVL-----GGGED 2220

Db 2182 SRTATATPSTKTRSTLLPSSPTAPITTVTTCGECQCAMSEMDYSPMGSGGDFD 2241

Qy 2221 -----GTF--PVSPEGKTEPGH--SRSAVYPLLYRQGBQ----- 2251

Db 2242 TYSNIRAGAVCEOPDLGCRAGQAPVPLRELQVVECLDPLGVCNRENOYQKFKMC 2301

Qy 2252 -----TSPSMGSKSPGNTSQPAPFESKLT-----ESNSAMVKSCKQ 2288

Db 2302 FNYEIRVFCNCYGHCPSPATSPSTPSPTGTTWILKLTATTTSTGSTATPSSTQ 2361

Qy 2289 -----EINKLNTNRENEEYNIQPGTEIFMMPAITGTGLMTYSOAVOEHAS 2337

Db 2362 GPAGTPHVTATPTPTTSKATP--FSSPG-----TATALPALSTATPTAT 2409

Qy 2338 TNNGLBAITRKALMGKYDQMBESPLSANAFNPLNASSLPAMPTAAGDSHT--L 2394

Db 2410 S--FTALPSSSLGTTWR-----LSQTTTPMATMTATSPSTPEV-----HTSTVL 2454

Qy 2395 TSPGGGAKXVSGRRSSKAKSPARGLASGDRPSPVS-----VHSEDCNRRPPLNRY 2449

Db 2455 TT-----TATTGATGATVATSPSTPGLAHHTKVTPTTTTGTTPSSPGRATP--PV 2506

Qy 2450 W---EDRPSAGSTPFP-----YNPLMR-----LQGVWASPPPGPLPAGSGP 2490

Db 2507 WISTTTPTTSGTVPSPISPGTHTPTVLTQTTPVATGSMATPSSSTQTSQTP 2562

RESULT 76

A35175

mucin 1 precursor, repetitive splice form A [validated] - human

N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistatlin

N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000

C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51

R:Lightberg, M.J.L.; Vos, H.L.; Gemmelsen, A.M.C.; Hilken, J.

J. Biol. Chem. 265, 5573-5578, 1990

A:Title: Epistatlin, a carcinoma-associated mucin, is generated by a polymorphic gene enc

A:Reference number: A35175; MUID:90202794; PMID:2318825

A:Accession: A35175

A:Molecule type: mRNA

A:Residues: 1-952,1033-1344 <LIG1>

A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PID:AAA5804.1; PID:g182124; GB:

A:Experimental source: splice form A

A>Note: Genbank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino-and carboxyl-er

A:Accession: B35175

A:Molecule type: mRNA

A:Residues: 1-19,29-952,1033-1344 <LIG2>

A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PID:AAA35806.1; PID:g182129; GB:

A:Experimental source: splice form B

A>Note: Genbank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino-and carboxyl-er

R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel

J. Biol. Chem. 265, 15286-15293, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithel

A:Reference number: A35886; MUID:90368715; PMID:1697589

A:Accession: A35886

A:Molecule type: mRNA

A:Status: not compared with conceptual translation

A:Residues: 1-19,29-992,1033-1344 <GEN>

A:Cross-references: GB:J05581; NID:g188869; PID:AAA59876.1; PID:g188870

A>Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence

R:Plan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716; PMID:2394722

A:Accession: A35887

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19,29-1109,1111-1339,1341-1344 <LAN>

A:Cross-references: GB:J05582; NID:g189598; PID:AAA60019.1; PID:g189599

A>Note: Genbank entry HUMPAMU contains four fewer copies of the tandemly repeated sequen

R:Wreschner, D.H.; Harevent, M.; Tsarfacy, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.

Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen CDNA sequences. Differential splicing may genera

A:Reference number: S10571; MUID:90276413; PMID:2351132

A:Accession: S10572

A:Molecule type: mRNA

A:Residues: 1-19,29-155,157-175,177-182,184-212,1033-1344 <WRE>

A:Cross-references: EMBL:X52229; NID:g37053; PID:CA36478.1; PID:g37054

R:Wreschner, D.H.

submitted to the EMBL Data Library, March 1990

A:Reference number: S40293

A:Accession: S40293

A:Molecule type: mRNA

A:Residues: 1-19,29-155,157-175,177-182,184-212,1033-1344 <WR2

A:Cross-references: EMBL:X52229; NID:g37053; PID:CA36478.1; PID:g37054

R:Abde, M.; Siddiqui, J.; Kule, D.

Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated

A:Reference number: A36735; MUID:9088473; PMID:2597151

A:Accession: A36735

A:Molecule type: mRNA

A:Residues: 1-142,144-162,164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:g181542; PID:AAA35757.1; PID:g181543

R:Maizawa, Y.; Miyauchi, T.; Hamano, M.; Ando, S.; Yoshida, J.; Takeo, S.; Shinazu, H

J. Biochem. 112, 609-615, 1992

A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu

A:Reference number: JX0235; MUID:93123189; PMID:1478919

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,1014-1017,1018-1032,1034-1037,1038-1057 <MAS>

A:Experimental source: gastric carcinoma cell

R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r

A:Contents: annotation

A:Reference number: S51026; MUID:95080414; PMID:7988707

A:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c

partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly

C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S46146

C:Genetics:

A:Gene: GDB:MUC1; PUM

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: 1621-1623

A:Introns: 20/1, 62/3, 1165/3, 1184/2, 1230/1, 1270/3, 1320/3

C:Superfamily: polymorphic epithelial mucin

C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphi

F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>

F:1-19,29-33/Domain: signal sequence #link PREB #status predicted <SIGB>

F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form

F:1133-1017/Region: 20-residue repeats (GSTAPAHGVTSAPDRPAP)

F:1133-1344/Region: mucin 1 carboxyl-terminal non-repetitive

F:1245-1272/Domain: transmembrane #status predicted <TM>

F:1046,1064,1118,1144,1222/binding site: carboxylate (Asn) (covalent) #status predicted

F:1213/binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 2.5%; Score 335; DB 1; Length 1344;

Best Local Similarity 21.5%; Pred. No. 0.0001;

1412 -----EGVATVKEAGRSIHIEIPREELHTELPRLAPRLKEGSIOTGTPLY 1459
DB 557 PGASAGASGGEGVSARPSOG-----TPSGTPASAPVAAGPAGG----- 599
QY 1460 DTGASTGSKKHDVRLIGS-----PGRTPPHPLDV-----MADARALE 1500
DB 600 GSGSGTSGS-----GSGAPAAARPGGGSDAPVDLITSGRGGSGAGNGAGTIGE 650
QY 1501 RACYEELKRRPRTASSGSGIARFAPVIVPELGRKROSPLTYEDHG-----APPAGHLP 1555
DB 651 TGAAGTGS-ASRPESGGTGGSGTGVASGKGTGGSGTGTGNGGAGAGVAGVA 709
QY 1556 RGSPTVMBEPTPRLGSLSSSKASQDKLITSP-----REIAKSPHSTVPEHHPI 1608
DB 710 RGG-----GGAGGAGNETRPLTCTVVVDVRLGKLKSAMTSFVMDASAGVQV 756
QY 1609 SPEHLLRGVS 1619
DB 757 WPDALIKGVA 767

RESULT 75
T45025
mucin MUC5B, tracheobronchial (imported) - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 21-Jul-2000
C/Accession: T45025
R/Designer: J. L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J. P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A/Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A/Reference number: Z2899; MIMD:9716151; PMID:9013550
A/Accession: T45025
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3570 <DES>
A/Cross-references: EMBL:Z72496; NID:G1834502; PIND:CA96577.1; PID:G1834503
A/Experimental source: placenta
C/Genetics:
A/Gene: MUC5B

Query Match 2.5%; Score 335.5; DB 2; Length 3570;
Best local Similarity 19.5%; Pred. No. 0.00028;
Matches 409; Conservative 203; Mismatches 835; Indels 649; Gaps 91;

QY 687 PAASEBAAPPVVEDEMAAGVSGNBEEMVEBAALHSGNEVIRGEGSGPATVNNSS 746
DB 824 PSSSTPGTTPPPLVLTATTPAATST-----VTPSSAL-----GTHHTPPVPTTA 870
QY 747 DTE-----SIPSPH-----TEAKDTGONGKPPATLGA-----DGRPP 780
DB 871 TTHGRSISPSPHVCTAMTSATGILGTHHTERSTGSHT-----PATTTGTHQSHPAL 927
QY 781 GPPTPRTSRAPTEP-----TPASEATGATPP-----PAPPSAPPPVPRK 824
DB 928 SSPHSRTTSPSPSGTTPGHTTATSRRTATATPKRTSTLLPQPT-SAPITTVT 966
QY 825 EKEKEETAAAPVVEEGEOKPRAEEL-----AVDTGKAEFVKSECTEAEKEGPAK 877
DB 987 MGEPPQCMSEMLDYSPMPGPGSGDDPTYSINRAAGAVCEQDLEGCQAQAPGVPLR 1046
QY 878 KDAEAETAGALKAEKKEG-----GSGRATTAKSGAPQSDSS 918
DB 1047 ELGGVVEGSLDFGLVGNREVGKFKOCFNYEIRVFCNNGHCHSPATSTATSPSTGP 1106
QY 919 ATCSADEVDEAEGDKNLLSPRSLITTPGDPRAANSPQKPLDKLQKRAAALPIQV 978
DB 1107 TTM-----ILTELTATTTSTGSTATPSTL-----KTAPPPVLT 1144
QY 979 TKVHEPPREDAAPTKAPAPPPPPONLQESDAPQPGSSPRGKSRPAPPADKEAPAAE 1038
DB 1145 TTATTP-----TWTSGATPSSSBGTATAP-----ALRSTATTTTATSVPIPISSSLGT----- 1194
QY 1039 AOKLPGDPPCWTSGLPFPVPPREVIKASPHAPDPSAF-----SYAPPGH 1082

DB 1195 -----TWTLSSQTTPTATMSTATPSSTETATHTSTVLATATTTGATGSAVTPS- 1244
QY 1083 PLPLGHDTPARPVL-----PPRPTISNPPILISSAKHNSVLEROIQAISQMSV- 1131
DB 1245 STPGTAHTTKVPTTTTGTTPATPSSSPGATLTPPVWISTTTPT-----TRKSIVT 1295
QY 1132 QLVHVPSEHAKAPVGPVMTWMLPLPMDPKLAPSGVQEOQLSPRQAGPESLGVPTAOE 1191
DB 1296 PSSIPGTHHTAVVLTITTTTVA-----TGSMAPTSSSTQISGTPSL---TTTA 1341
QY 1192 ASVLGRTALGSPGSGITTKGI-PTRVPSDAITYRGSITHTGPADVLYKGTITRIIGED 1250
DB 1342 TTI-----TATGTTNPSSGTPGRPIPVLTATTPAATSTVTPSSAL---GT-THRPVP 1395
QY 1251 SPSRLDNGREDLSLPGHVIYEGKGNHLVYEGGMSVYQCKEGRSSGPHETAAAPKT 1310
DB 1396 NTATYTHGRSLSPSPPTV---RTAMTSATSGTGLGTHITE---PSTGSHPAATGT 1448
QY 1311 YDMMEGRVGRASISASIEGLMGRALP-PERHSPHNLKEOHNRISITQGI PRSVVEAOED 1369
DB 1449 TQ-----HSTPLSSPHSPSRSTESPPSGTTPGH-----TTATSRTATATPS 1493
QY 1370 YLRBAKLKKEGTPPPPPRDLTEA-----YKQALPPLKAPKPHEG 1413
DB 1494 KTRTSTLL-----PSSPTSAPIYTTVWGCSEPCAMSEMLDYSMPGP---SGDDFD 1543
QY 1414 LVATVKEAGRSIHIEIPREELHTELPRLAPR-----PLKEGSIOTGTPLYDTGASTT 1466
DB 1544 TYSNTRAAQGANVCOP-----LGLECAQAPGVPLRE-LGGVVEGSLDFGLVCR 1592
QY 1467 GSKKHVDVSLIGSPGRTPPPVPLDVMDARALRACYESLK-----SRPGTAS 1516
DB 1593 N-----REQYCK-----FKMCFNYEIRVFCNNGHCHSPATPSS 1626
QY 1517 SSGSANGAPVYVELKRPQSPPLYTDHAPRA---GHLPGRSPVMBEPTPRLGEGSL 1574
DB 1627 TATPSTPGTWWLLEQTTAATTAATTTGTAIPSTPETAAPPKVILISQATTP-----TA 1681
QY 1575 SSSKA-----SOPKLTSPREIAKSPHSTVPRHHPIPSYEHLLRGVSGVDLYSHIPLA 1631
DB 1682 TSSKATSSSPRTATLLVLNSTATKSTATGTPPSSSTL-----GTTGQSNBPHDMA 1736
QY 1632 FDPSTIPRGIPDAAAAYVLRHLAPNPTVPHLYPIRLGYPPRALLENQTIINDVIT 1691
DB 1737 TWST-----IHPSST-----PETT----- 1750
QY 1692 SQOMHNTATMAORADMRLGLSPRESSIALNVAGPRGIIDLSQVPHLPVLPPTGTP 1751
DB 1751 -----HISTVLTATTTTATSSMSTPS-----TPGTT 1779
QY 1752 -----ATMDRLAYLPTAPOPSSRHSSSPSPGCPHNLKPTTSSSEKRRDRDRD 1805
DB 1780 WILTELTATTTAALP-----HGTPSTPPTTWILTEPSTTAVTVPTGSTATAS 1830
QY 1806 RDREREKS---ILSTTVYHAIWPRGTEQSGSGSGSGSGSGSSSPASHAHQHSPI 1863
DB 1831 STRATAGLKVLITATTP-----TVISRAAPSS-SEPTATALALAKTA----- 1875
QY 1864 SPRTDALQORPSVLNMGKGIITAVEPSKPTVLKSTSTSSPVPAATFPATPCPLQG 1923
DB 1876 TPTTANSVTAIPSSSLGTMTLSQTTT-----TATMSTATPSSGT-PEVATSTVTL 1926
QY 1924 TLGQVPTIMEPVLLPKAPRVARPERPRADTGHAFALKAPRARG-----LBPASRP 1975
DB 1927 TTAT--TTRGSAVATPSSSTPGTAHTTKVPTTTTGTATPSSSGTALTTPVWISTTTT 1985
QY 1976 -SKGSEPRPLVPVSGH-ATIAKRAKNLARNHNSDPRAPASADPRREKTOGKSPFSI 2033
DB 1986 TRGSGTVTBSLPGTHHATVLTITTTTVAATGSMALPSSSTQSGRPBLYTTAT----- 2040
QY 2034 QLELRSILGYHSGSSYPE---GVEPVSPV-----SSPLTHDKGLPKHLELDKSHLGE 2085


```

QY 817 -APPVVPKEKEBETAAAPVEBEE-----QKPPAAEELAVDTGKAEEVYKSECT 867
Db 221 GABGF-----GQFQGNPGEPSGAGCMGPRGPGPGKGDDEAGKPKGS--- 268
QY 868 EBAERBPAAKCAEAALAT-----ABGALKAEKKGSGSRATTAASGAPQDSASACS 922
Db 269 --GEGRPBPQAGRGPGTGLPGVKGHRGYPGLDGKGEGAA--GAKGSGASGAG 323
QY 923 ADEVEBAEGDKNRLLSPRSLTLTGTDRANASPOKPLDKLKORAAATPPIQVTKVH 982
Db 324 A-----PGMGRGLPGERG----- 338
QY 983 EPPREDAPTK-----PAPRAPPONTLPESDAPQPS-----SRPG 1021
Db 339 RCGSSGAGARGNDGLPGAGPPGPGVGPAGAGPAGPAGSGEAGPTGARGPAGAGP 398
QY 1022 KSRSPAPPADKFAFAEAOQLGDDPPCWTSGLPFPVPREVIKASPHADPAGAFYAP-- 1079
Db 399 ESGTSGSPSGA-----SGNP--GTDGLP-----GAKGSGASGAGIAGBGF 438
QY 1080 PGHPLPLGLHDTAPVLPPTTISNPPLISSAKHPVLERQIGALISQMSVOLHVYSE 1139
Db 439 PGRGPGPGQAGTGLPGKGT-----GDPGLAGPKG-----HGPKGE 477
QY 1140 HAKAPVGVYTMGLPLM-----DPKCLAPSGVQKOBOLSPRGQAGPESLGVPTA 1189
Db 478 IGSA--GP--QGAPEGAGEGKRGARBEAGAP-----LGPGERGAGPQNRGPF 525
QY 1190 QEASVLRGT-----ALGSPGSGITTKGIP--STRVPSDAITVRSGITHTPADVLYK 1241
Db 526 DGLAGRGKARGENGVLGPKKGNDDPRGPRGLRGRLT-----GRRGAGP 578
QY 1242 TI--TRIGED-----SPSRLDGRBDSL--PGHVIYBEGKCHVLYSGMSVTC 1289
Db 579 KVGPSSGAGBDGRRPGRPOGA--RQPGVWGFPGRGANGEPKAGE--KGLGAPGLRGL 636
QY 1290 SKEDGSSSGPRHETAP-----KRTYDMMEGRGR-----AISAS1 1327
Db 637 PGDGETGAGGPNRGPAGEGEGEOPRPGSGFGLRGPRGSGEGEGKGDQVPEBA 696
QY 1328 EGLMG-----RAIPRERHPNHLKEONHINGSITOGIPRSVVEAOEDYLREAKLIRREGT 1383
Db 697 AGLVGRGERGPFGRERS-----SGPGLQGS--RGLP-----GT 729
QY 1384 PPPPPSRDLTEAVKTOALRPLKRAHEGLVATVKEAGSITHEIRREBELRHTPELAP 1443
Db 730 PGTDFPKGATGSPGNGAGP-----PGLQMPGERGAAGIS--GPKGMDGTGEGKPEG 782
QY 1444 RPLKESITGTPLKVTGASTTSGKNDVRSLLGSGRTFPPVHPLDVADARALERAC 1503
Db 783 APGKDS-----RGLTGP-----LGRPG-----PAGP----- 804
QY 1504 YEESLKSRPGTASSSGSIARGAVIYBELGKPRQSPLYEDHAPAGHLPRGSPVTR 1563
Db 805 --NGEKESGPGSPGIVGARAGADRGENGPP--GP-----AGPAG--PPGS----- 846
QY 1564 EPTPLRLOEGSISSSKAODKLTSTPREIAKSHISTYREHHPRLISPYEHLILAGVSGVD 1623
Db 847 -----DQO--AGLKGDQGE--SGQKGDAGAPRQSPSAPRQGP-----TGVMG-- 887
QY 1624 YRSHIPLAFPTSPRGIPGLDAAAYLIRHLAHPNTPYRHLIPYRLLRGVPTDALLENRO 1683
Db 888 -----PKAGAGAGPAGATGPRGA--GAVGTGPR--GNRGP-----GPPGSAKGKGP 934
QY 1684 TIINDYTSQGM-----HNTATAMAGRAM-----LRGISPSSSIALNYA 1726
Db 935 GVRDAGBTGRAGDPLQSGPAGAGEKEGPEDEGSPDPGSPQGLSGNRGIVGLRQ 994
QY 1727 GPRGIDLSQVPH--PVLVPTPTGTATAMRLAVALPTAPQFSSSHSSPLSPGPT 1784
Db 995 GERG-----PGLRPSGEPGKQGGPSSGDR-----GPRPVPFPG 1031
QY 1785 LTKPTTSSERERDRDRERDRERKSLTSTTV-----EHAPIWRPTEOSSGSSG 1839

```

```

Db 1032 LTGPAGTGRGKNGSPGPPGRD-----GATGKEDRGHTGPIGAGAGSPGAPG 1082
QY 1840 SSGGGGSSSRPASHA-----HOHSDISPTODA-----LQRPVSLAN--TG 1882
Db 1083 PVGPTRKQDGRGEGSGPLGPGSGPAGARGLNGPGRGKGEANGEGEKQKHRTGT 1142
QY 1883 MKGIITAVEBSPKIVL--ASTSTSPVRPAATPTPTHTCPLGGLDGVVPTIMEPVLLP 1939
Db 1143 LQGL-----PGPPTAGDQAGTSAGSPGPRG--PPGPVPSG--KDG-----S 1182
QY 1940 KEAPRVARPERPRADTGAFLAKPARSGLEPASSPSGSEPRELVPSGHATIAPTA 1999
Db 1183 NGTGPICPPRPRRSGETGAGPAGSG-----PPGPRGPRG--IDWSAF 1228
QY 2000 KNLAPHHASDP-----PAPPASAS 2019
Db 1229 AGLSQPEKGPDMRMYRADQASSS 1252

```

RESULT 73

```

126517
Hypothetical protein Y18D10A.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26517
R:Haftis, B.
Submitted to the EMBL Data Library, December 1998
A:Reference number: Z20226
A:Accession: T26517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1634 <MIL>
A:Cross-references: EMBL:AL034393; PDB:CAA22308.1; CESP:Y18D10A.1
A:Experimental source: clone Y18D10A
C:Genetics:
A:Gene: CESP:Y18D10A.1
A:introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

```

Query Match 2.6%; Score 337; DB 2; Length 1634;

Best Local Similarity 19.7%; Pred No. 0.00011;

Matches 295; Conservative 181; Mismatches 530; Indels 494; Gaps 64;

```

QY 172 LIQNDRV-----DREITVWEQISLK-----KQOQLEBAKPEPEKVPSPPT 219
Db 26 LMQNGQRAEMQDFKQVQBECEVLMKLVSTDAVRESVQREBAEAINAK----- 74
QY 220 ESKHRSIVQIYDENRKAABAHRLLEGIGVVELPLYNQSPDTPQYHENTKINQAMRK 279
Db 75 ETAEASLKQGYDSEFYKKSA-----SLQEQDDAEKHKQ----- 111
QY 280 LILYFRNRHAKQKQKFCQRYDQIMELTEKVERIENPRRAKESKVEEYKOPPE 339
Db 112 -----ANYRANK-----YDSVC-AL--TRLEANN-----ID 135
QY 340 IRKRELQRMQSRVQSGSGL-----SMAARSH-----EVEIINDGLSEQ 382
Db 136 IKSKESTYKLETKETKMTSLDNOVTLTKAKKINSSKTKLDYLYKSVIQVDALST 195
QY 383 ENLEKQRLAVIIPMLYADQQRKIFINMGIMADPMKVYKQVQVMMWSHQBETPRE 442
Db 196 SNIKQSKSI-----KDLTVFR-----RDDVRL 219
QY 443 KFWQHPKNGLIASFLEKTVACVLYYLTKKNENYSLVRSYRRGKSQOQOQOQO 502
Db 220 FNSQ-----WKSLSTSPSISATEASESDEPE 249
QY 503 QOQOQOQOQMP-----RSSQEKQEKQKAEKEENP--EVENDKEDILKEXT----- 550
Db 250 HLDVITTEBAPRIGEQTLSDDEEBEVEDADEVREIRQLNEELLQRYKEXIVRR 309
QY 551 DQSGDNDKEKAVASGKRTANSQGRRKRI--TRSMANEANSSEALTPQOASIASME 608

```

```

Db      7934 GIPLEPTDA-----SGA---VIGPDGELIIPDASGKPLSADGSPLEPTDNNGNVYLVPADE 7984
Qy      1208 ITKGIPISTRVPSDAIYRGSIITGTCPADVLYKGTITRIIGEDSPSLDRGRSDSLPGH 1267
Db      7985 VTTKV---LPTDSD---GNVYH-----PIRPGD--TFLGTDASGSFVTDGQ 8024
Qy      1268 VIVECKKHVLSYEGSMVTOCSKSDGRSSGPHH---TAAPKTYMMGKRGVRAIS- 1323
Db      8025 AIEKODEKPIGPDQVLPDASGNVYIPVIGPDQALPTDCKGKTVVPVPGDPTPLST 8084
Qy      1324 --SASIEGLMGRAIPERHSPHLLKEQHIRGSIITQGIIPRSYEAQEDYLREALLKRE 1381
Db      8085 DASGLILGDEPIPTD-----ASGKPLSAD 8110
Qy      1382 GTPPPPPSRDLTEAKYTOALGPLLKAHGLVAT--VKEGSIHEI PREELRHTEPL 1439
Db      8111 GSPLP---TDAVGNVYILVPSDDGVIRTHPTDESGNTIYITPKD----- 8151
Qy      1440 PLAPRPLEKSGITQCTPLKYD--TGASTTGS---KKHVRSLIGSPGRTPPP-----VH 1488
Db      8152 -----GTPLATDSTGAFTVDDGVITKDEKPIGPDQVLPDASGNVYI 8197
Qy      1489 PLDVWADARAL-----ERACEEESLKSRPGTAS----- 1516
Db      8198 PV-TSSDQVLPDIAKPIVVDGSKPLPTDASGNVYIDNNGKPIYIEGEEPTGPDQKLS 8256
Qy      1517 -----SSGGSIAAR-----GAPVIVPEL 1533
Db      8257 KNEKEGWYPLVDKFGKFEVETDNDKPVLYTVVNDNGNELSKNDGNMIDLSENEIDTDEL 8316
Qy      1534 GKPROS---PLTYEDHG-----APFAGHLPRGSPVMREBTPRLQGS 1573
Db      8317 GRPLDSGPNPKFNDGNHVAIPAQIEEBETTPAIFIT--IIDGPIEDDDVYTDKGN 8374
Qy      1574 LSSSKA-----SQDRKLSTPREIAKSPHSTV--PEHHPIPSIPEHLRLG 1617
Db      8375 VIPTNSEKPIDENGQVLPKNEDEPFVK-PKS-ADTTGSTIYSPGSPLEPTD----- 8424
Qy      1618 VSGVDLYNSHIFPLAFDPTSIP-----KGIPLDAAANY-----LPHHLAPNPTY 1661
Db      8425 ASGAATIGDGEPIPTDSSGRPIPSKDGSPLPTDASGNVYILVPSGEGVTDLSPTDEAGNTIY 8484
Qy      1662 PHLYPPIYLRGPDAAL--ENKQIINDYITSQGMHNTAAMAQRADMRLGSLPRSS 1719
Db      8485 PITKPDGTLATDSTGTFVADGQIIEKD-----DEKPI----- 8519
Qy      1720 LALNYAAGPRGIT--DLSQVPHLPLVLPPTGPTATMDRLAYLPTAPQ--PFSSRHSS 1774
Db      8520 -----GPDQVLPDASGNVYIVIGPDQALPTDSSGKTVVVRGPDGTPRLPTDASG 8572
Qy      1775 SPLSFGG--PTHLT-KPTTSSSERERDRER--DRDREREKSILTSTT-TVEHAPI 1826
Db      8573 AVTIGDGPPIPTDPSGKPLSADGSKLPTDINGNVYLVPADEVTTKVLPTDSDGNVYH-PI 8631
Qy      1827 WRP-GTEGSSSGSSGSGGSSSRPASHSHAHQSPISPRQDMLQGRPSVLANHTWKMG 1885
Db      8632 TRPDSTPLGTDASG-----SFTTEGQIVKNDG--KP-----TGPDG 8668
Qy      1886 IITAVEPSKPTVLARSTSTSSPVRPATATPPTATCPAGTLDGVYPLTMEPVLLPKPEARV 1945
Db      8669 QVLPTDSDSNVYIYPIGSDQAMPPTDT-----TGSVITYLVPSDGVIGISPPKV 8717
Qy      1946 ARPERPRADTGHAFLAKPPRASGLERPASSPKSGEPRPLVPPVSGHATITARPAAKNLAPH 2005
Db      8718 AKP-----VGPDKVLPDASG-----H 8735
Qy      2006 HASPPRPAASDPNHRKTKSQKPSIQELRLSLGYHGSYS-----PEGVEPVSPV 2060
Db      8736 FIGPGPIPTD-----YGVTVYSDTVTTTPDGI-----PUS 8764
Qy      2061 SPSS---LTHDKGLPKHLEELDKSHLGEELRPK-----QPGFVKLGGEAA 2101

```

```

Db      8765 NDSTGAFITEDGTIENNEDGKPIGPDQVLPDAGVNYIYPAIGPDQALPTDESGNV 8824
Qy      2102 HL---PHLRPLRES-----QPSSSLQOTAPGV-----KHQRVYTLAQH 2138
Db      8825 YVPRGPDGTPPLTDVSGAVIGPDDEPIPTDASGKPLSADGSPLEPTDNNGNVYLVPADEV 8884
Qy      2139 ISEVITQDYTHHHPOQLSAPLAPL-----YSF--PGASCPVLDLRP--PSDLVYLP 2187
Db      8885 TTKVLPTDSDGNVYHPIITRPDGTPLGTDASGSFVRDDQALAEKDEKPIGPDQVLPD 8944
Qy      2188 -----PDHGAPARGSPHSEGGKRSPEPKNTSVLGGEGDIEFVSPPEGKTEFGHSRSA 2240
Db      8945 ASGNVYIPVIGPDQALPTDESGKT-----YVPRGPDGTPPLTDASGA 8988
Qy      2241 VYPLLYRDEGTEESRNGSKSPGNTSQPRAFFSKLTESNANMYSKKOEINK-LNTHNR 2299
Db      8989 VIGL--DGEPIPTDASGKPLSAGSPLE-----TDNNGVYLVPADEVTTKVLPTDSD 9039
Qy      2300 NEPEYNISQPGTEIFNMPAITGTGLMTYRS-----QAVQEHASTMGLAIIIRKALM 2351
Db      9040 GNVYHPIITP-----DGTPLGTDASGSFVRDDQALAEK----- 9073
Qy      2352 GKTDQWESPPLSANAFNPPLNASASLPAAPIYAADGRSDHTLTSPGGGGAKY-----S 2406
Db      9074 -----DEGKPIGPDG--QVLPTDASGNVYIPVIGPDQALPTDES---GKTVVPVRGPD 9122
Qy      2407 GRPSSRKAKSPAPPLAGSDRRPSSVSVAHSEBDCNRRTPLTRWEDRPSSAGSTPFPVN- 2465
Db      9123 GTPLEPTDASGAVIGL--DGEPIPTDAS-----GKPLSAGSPLPTDN 9162
Qy      2466 -----PLMRLOAGVMAASP--PPPLPAG--SGPLA--GPHHAWDEEPK 2503
Db      9163 NGNVYLVPAHNVTTKVLPTDSDGNVYHPIITRPDGTPLGTDASGSFVRDDQALAEKDEGK 9222
Qy      2504 PL 2505
Db      9223 PI 9224

```

```

RESULT 72
A0333
collagen alpha 1(I) chain precursor - African clawed frog
C.Species: Xenopus laevis (African clawed frog)
C.Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C.Accession: A40333
R.Su. M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A.Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A.Reference number: A40333; MUID:92011898; PMID:1918153
A.Accession: A40333
A>Status: nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-1492 <SUA>
A.Cross-references: GB:M63596
A.Note: this sequence is presented as substitutions relative to another sequence in a fi
e they replace: the appropriate interpretation of the sequence figure was reconstructed
C.Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C.Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F.37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F.1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

```

Query Match      2.6%  Score 337; DB 2; Length 1492;
Best Local Similarity 22.3%  Pred. No. 9.6e-05;
Matches 327; Conservative 99; Mismatches 556; Indels 482; Gaps 74;

Qy      726 ASGNEVPRGEGSGPATVNNSSDTEISPEPATEAK-----DTGQNGPK---PP 770
Db      101 SSGGQVAKGGEPRDIDVVGPRGPFGQPSGEGPRGEGDKGTGAGPRGRDGE 160
Qy      771 ATLGADEP--PPGPTPRRTSRAPIEPTPA--SEATCA-----PPPPAPPS 816
Db      161 GTPNGPAGPGRPGPGGLGAGNFAAQWGTGFDERKAGAGQNGVMOGPMGPGPRGPGPS 220

```


Db 1394 PPKSKSLRA-----DTEBEFLAFKKQPPSAGKANH-----TP 1426
 Oy 1444 RPL-----KEGSIOTGPTL-KYDTGASTTGSKX-----DVSRLIGSPGT 1483
 Db 1427 KAVGEKNOINFLGTPVQKLOPGLPGSNRLQTRKEKAQALELNTFFELFQTPCTD 1486
 Oy 1484 FPPV-----HPLDWADA-----RALEKCYEESLSKSRGTASSGGGSLA 1523
 Db 1487 NEPLADEKTKILLCKSPQSDPADTPNTKQRPKRSLKKADVEEFLAFKLLPSAGKANH 1546
 Oy 1524 KGAAPVVR-----LGKROSPLTVEHDGAPPAHLLPRGSPVTVREPTRLQEGSLSS 1577
 Db 1547 TPKAAVGEKNDINTVGTVEK-----LDLGNLP-GS--KRRPQTPK-----EKA 1589
 Oy 1578 KASQD-----RKLSTP-----REIA-KSPHASTVPHHHPISPYEHLRGVSG 1620
 Db 1590 KALEDLAGKRELFGTGHTEESWTDKITEVSCSPQ---PDPVKTPSSKQRLKSLGK 1646
 Oy 1621 VDLYKSHITLA-FDPTSIIRGLPLDAAAYLPRHLAPPTYPHLYPLILGYPDTAL 1679
 Db 1647 VGKKEVLHPVGLTOTS-----GKTYOTH----- 1670
 Oy 1680 ENKQOTIINITYTSQOMHNHTATAMAQADMLRGL-----SPRESSIALNYAAGPRGIDL 1734
 Db 1671 --RETAGDG--KSIKAFKESAKOMLDPAVYGTGEMERPTPKKEAOSLEDLAGFK---EL 1723
 Oy 1735 SQVP-HLPLYVPTPCTPATMDRLAYLPTAPQFESSRHSSEFLSGFTHLTKPTTSS 1793
 Db 1724 FQPPDHT-----EESITDDKTKIACKSP-----DESMDTPST-- 1758
 Oy 1794 SESEPRDRDRDREREKSI---LTSTTVHAP-----IMRPTGEOSSGSSGSG 1842
 Db 1759 --RRREKTPYKGDIVEELSAKQILQTHTDVDEDEKGINVEFKTAQKLDPAASYT 1816
 Oy 1843 GGGGSSSRPASHGHAHNSPISERTODALQORPSVLANTGMKGI--TAVEPSKPTVLRST 1901
 Db 1817 G-----SKQPRTRKGAQRLLEDL--AGKELFQTPCTDKPPTTHK 1857
 Oy 1902 STSSPVRAATPPRATHCPGLGTLGVPTLMEP-----VLPKAPRYA 1946
 Db 1858 -TKIAGSPQDRPVGT-----PTFFPQSKRLRKADVEEESIALRKRPSPVG 1904
 Oy 1947 RP-ERRPADTG-----HAFIAPRASGLEPASPCKSGSEPRPLV-----PVSQHA 1992
 Db 1905 KAMDTPRAGGDEKMDKAFMGTPVQKLDL--PGLVP--GSKWFPQTPKKAQALEDLAGFK 1961
 Oy 1993 TIARTPAKN-----LAPHNASPPRAPPASASDPHREKTOGKPPSIOE--LELRS 2040
 Db 1962 ELQGTGTDPKPTTDEKTKIACKSPQDPVDTPASTK--QRPKNLRKADVEEFLALR- 2018
 Oy 2041 LGYHSSSYSPGVEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRPKQPGVYKLGGA 2100
 Db 2019 -----KRTPSAGKAMD-TPKPAVSDKINTEVET-----FVQKLDLL 2055
 Oy 2101 AHLPHLRPLRESPSSPILQTAQVYK-----GHQKRVTLAQHISEVITOD-----Y 2147
 Db 2056 GNLPGRKOPQTPKKAQALEDLVGFKEFLFQTPGHTEESWTDKITEVSCSKSPQSPFK 2115
 Oy 2148 TRHHPOOLASAPLAPLSPGASCPVLLRRPP--SDLYLPPPD-----HGAPARSPHS 2200
 Db 2116 SRSSKQRLKPLVK-----VDMKEEPLAVSKLRTSGETQTHTEPTGDSKSI 2163
 Oy 2201 EGGKRSSE---PNKTSVLGGGED-----GIEP-VSPPEGMTEGHSRSAVYPLVR 2247
 Db 2164 KAFKESKQKQILDPAAASYGSRRLQTRKEKARALEDLVDFKELFSAGHTEESM----- 2217
 Oy 2248 DGBETESRMSKSPGNTSQPAPFSLKTESNSAMVSKKQOINKKLNTRNREPEYNIS 2307
 Db 2218 --TIDKNTKIPCSB-----PPELITPATSTKCPKTRPKVEYEBLSAVERLL---QTS 2267
 Oy 2308 QPGEIENMDALITGTGLMTYRSQA-----VOEHAS----- 2337

Db 2268 GOSTHTHKEPASGDEGKIVLKQARAKKPNVEEESRRRPRAPKKAQPLEDLAGFTELS 2327
 Oy 2338 -----TNMGLFAIRKXLMGYDQWE--ESPPLSANANPLINASASL----- 2377
 Db 2338 ETSQHTQESLTA-----GKATXIPCESPLEY-----VDTTASTGRHLTRTQOKQV 2375
 Oy 2378 --PAMPTTADGRSDHTLTPSGGGRKAVSGRPSRKAASPAQLAGSDRRP 2428
 Db 2376 EEPKAVKTYQTSGETTADKKEPAGEDKGYALKKESAKQTPAPASVYGSRRRP 2428
 RESULT 71
 T16580
 hypochelical protein K07E12.1 - Caenorhabditis elegans
 C1Species: Caenorhabditis elegans
 C1Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C1Accession: T16580
 R1Fulton, L.
 submitted to the EMBL Data Library, May 1994
 A1Description: The sequence of C. elegans cosmid K07E12.
 A1Reference number: Z18540
 A1Accession: T16580
 A1Status: preliminary; translated from GB/EMBL/DBJ
 A1Molecule type: DNA
 A1Residues: 1-13055 <FULL>
 A1Cross-references: EMBL:U00054; NID:G485140; PID:G485141; PIDN:AAA50715.1; CESP:K07E12.
 A1Experimental source: strain B16col N2
 C1Genetics:
 A1Gene: CESP:K07E12.1
 A1Intron: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3;
 ; 6014/3; 6159/3; 6665/2; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1229
 Query Match 2.6%; Score 338; DB 2; Length 13055;
 Best local similarity 19.3%; Pred. No. 0.00092;
 Matches 417; Conservative 239; Mismatches 762; Indels 744; Gaps 106;
 Oy 674 EKENARRKKKKAAPAAESEAAPPVVEDEMEASGV--SGNEBEVEEADALHASGNE 730
 Db 7477 EDQKLSKKKGWVYPLVDKFGKPEVETDNDKPVITVVDNDGNELSKNDGKWDLSGNE 7536
 Oy 731 VERGECGPATYNNSSDTESTP-----SPHTEAKDTCQNGPKPATLIGADGPPGPPPT 785
 Db 7537 IDTDLIGRPL-----DEGMPYKFDNGHVIAQIEEBEATPA----- 7576
 Oy 786 PARTSRAP--IEPPASEATGAPPPAP--PSPAPPPVVPKKEKEETAAAPVVEG 840
 Db 7577 -----IPPIITIDGPIINEDGVYTDKQGNVITPNSSEKPI-----DENQGVLPKNGYG 7624
 Oy 841 EEQKPPAAEELAVDTGKAE--EPVSECTEAEBSGPAKDAEAAEATAGALKAKKEG 898
 Db 7625 EFVKPKEA-----DTTQSTIVSPDGSPLPTDA-SGAAIGPGEPIPTDSSG--RPLSKDG 7676
 Oy 899 G-----SGRATT-----ASSGAPQSDSASATCSADEVDEA 929
 Db 7677 SPLPTDASGNVILVDSGCVTSLPTDEAGNIYITPDTGLATDSTGSLVADGQIT 7736
 Oy 930 EGGDKNRLSPRSLSLTPGDPRANASPOKPLDKQKORAAAIPIQVTKVHEPREDA 989
 Db 7737 EKDDKEKPIGPQGVLP-PTD--ASGNVITP---IGPDGALPTDSSGKTVYVVR--- 7785
 Oy 990 APTKAPAPAPPON-----LOPESD-APQOGSSPRGKNSPAP-----PAKKE 1033
 Db 7786 -----GPDGTPPLPTDASGAVIGPDEPIPTDPSGKPLSDSGPLPTDINGNVILVPSDES 7840
 Oy 1034 APAAEQKLRGDPPCMTSLPRPVPRVETKASPHAPPSASVAPRPGHPLPLGLHTAR 1093
 Db 7841 A-----AKVLPTE-----SGSVVPIPKPD--CTPLCTDASGSFVTDGQ--AIGDDDEK 7888
 Oy 1094 PVLPRPTISNPPLISAKHPSVLEROIQAISQGMVQLAHVYSEHAKA--PV-GPVTM 1150
 Db 7889 PIGPDGQTL-----PIDDSGNY--IYVVGPGQGL-----PTDESGLTVVPLGP--D 7933
 Oy 1151 GLPLPMDPKKLAPFGSGVKOEQLSPRGQAGPPESSLGVPFPAQEAASVLRGTRLSG---VPGGS 1207

2214 VLGGEGDIE-PVSPBEGMTBPG-----HSRS-----AVY 2242
1436 ----GDDYPCGPGPGPIGDGPGKPGPGYLCGLVLVHSOTDPEPCPLMPRLMTGY 1491
2243 PLVYRGDEQTEPSR-----MGSKSPGNTQPPAPFESKLITSNASAMVSKKOEINKKANTHN 2298
1492 SLVLYEGEKHANNQDLGLAGSLPVPFSTLPAYCN-----IHQCHYAO 1535
2299 RNEPRYNISQPGTEIFENMPAITGTGLMTYRSQAOVEHASTMGLFATIRKALMGKYDWE 2358
1536 RNDSSYMLASAP-----LP-----MPLSEBAPRIPYRCAVCAPQAQAAVAHSDQ-- 1583
2359 ESPPLSANAFNPLNASLPAAMPITADGRSDHTLTSPG 2398
1584 -STPPCQPTWRSWLTIGYSF--LMTGADQGGGQALMSPG 1620

RESULT 70
B48666
cell proliferation antigen Ki-67, short form - human
C/Species: Homo sapiens (man)
C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C/Accession: B48666
R/Schlueter, C.; Duchrow, M.; Mohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A/Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ous.
A/Reference number: B48666; MUID:94043435; PMID:8227122
A/Accession: B48666
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2897 <SCH>
A/Cross-references: EMBL:X65551
C/Superfamily: kinase interaction domain homology
C/Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F,29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.6%; Score 338.5; DB 2; Length 2897;
Best Local Similarity 18.7%; Pred. No. 0.00017;
Matches 527; Conservative 348; Mismatches 1107; Indels 831; Gaps 131;

24 LSPVQIARHTDVLLEVOHNSRDYASHLSP-----GSII-OPORRPS-----LLSEFO 73
39 IQLPV-VKQKCKIIEHQBALHNFSS-TNPTQNGSVIDEIPVULKAGDVITITIDRSFR 96
74 PGNERSQELHLPRESHSYLPBLGKSEMEFIESKRPRLBLDPDLLRPSPLATQGPAGE 133
97 YENESLQN-----GRKSTEP-----PRKIREQEPARVSRSSPSSDPDEKA 137
134 DLTK-----DRSLTGKLEPVSPPSPHPTDPELVLPRLSKEELIQMMDRDRDEITWE 187
138 QDSKAVSKITBEGKSGNPELFDENLPNTPLRKGAPTK--RKSIVMHTPPVLKTIIEQ 195
188 QOISLKKKQOOLEE-----AAKPREPEKVPSPPIESKHSLSVQIIVYDENRKAENA 241
196 PQPSKQSGSGSEIHENVAQSLVISPAPSPKKTIVASDQRRSKTAPASSKSQTEVP 255
242 HRILEGLPOVELPLYNQPSDTRQYHENIKINQAMRKULIYFKRNHA-----RKQW 294
256 KR-----GGERVATCLOKRVISIRSQHDILQ-----MICKSRGSGSEANLIVAKSW 302
295 KQKFCQRDQMLEALEKVERIENNPRRAKSKATREYIEKQPPRIKROREIQEMQSRV 354
303 ADVV-----KLGAQOTQTKVTKHGPOR-----SMNKRQRRPATPKKPV 340
355 GORGSGLSMSAARSHEVSEIIDGLSEQENLEKQKQOLAVIPPMIYDADQOQIRKINMG 414
341 GEVHQSFTGHANSP-----CTIITGKAHTEKVVAPRPRIVLANPL-----SNQMDKEDLS 394
415 LMADPKYKQKQV-----NMMSQEK--ETPREKMQHQRKNGGLIASFLE 459
395 GIAEMFKTPVKEQPOLTSTCHIAISNSEYLLGKQFGQTDGSEBPILLPTSESFGNVFSA 454

460 R-----KTVACVLVYUUTKKNENYKSLVRSYRRRGKSQOQOQOQOQOQOQOQ 511
455 QNAKQPSDKCSAPPLPRROCIRENGVAKTPRTTYKTSLE----- 496
512 MPSSQEKDEKEKEKEKEKEEKEEVENDEKDLKKTDDTSGEDNDEKAVASKGRT 571
497 -TKTSDTEBESKSVTVNNSGRSTEFPRNIQKLPVESKSEBTNE-----IYECILKRGKA 552
572 ANSGRRKGRITRSMANEANSEALITPOQASLASMELNESKRTTEEMETAKKGLLEHG 631
553 TLLQORRREGM-----KEIERPFEYTK-ENIELKEN-----DEMKMKR-----S 592
632 RNV-----SARVMGKTVSQCKNFYFNYYKRONLDELLOCHKMEKERRARAKKKAP 687
593 RTWQCKCAPMSDLDLKSIPDTE-LMDTARGOULLQ-TQDHAAKPSKEG--KITMP 647
688 AAASE-EAAPPVVEDEMEANS-GVSGNEEMVEBAALHASG-----NEVREGSCGPA 740
648 CQSLQPEPIINTPTHTKQOLKASLGKGVKEBLAVGFTRTSGETHTHREPAGDKSIR 707
741 TVNNS-----SDTESI-----PSHTAAAD--TQO 764
708 TFKESPKQILDPAARVGMKMPRTPKBEAQSLDLAGFKELFOTPGSESMTEDETKT 767
765 NGPRPATLGADGPPGPPRPRRTSRAPIR-----TPASEATGAPTPPAPSPS 816
768 IACKSPPESDVDTTSTKQPFKSLRKADVEEFLALRKLTLP-SAGRAMLTPKA----- 821
817 APPVVPKEKEKEBETAAPVE-----EGEBQKPPAEELA-----VDT 855
822 -----GGDEKDIAFMGTPIVQKLDLAGTLPGSKRQLOTPKKAQALDLAGFKELFQTP 875
856 GKABEVP-----KSEC-----TEABEBGPAKCD--AEABATBGALEKAKK--EGGS 900
876 GHTEELVAAGKTTIKPCDPSQSDVDVPTSTQKQPKSIRADVEGELLACRNIMPASGK 935
901 GRATTAASSGAPQSDSDSATCSADENDEAEG--GDQKRLSPRSLLTPPTGDPANASPO 958
936 AMHTPKRSVGEKIIIFVGTIPVOKLDTLEULTSKR-----PQ 975
959 KPLDLKQIKQRAAIPPIQVTKVAEPREDAAPTKPAAPAPPPONTQPSDAPQPGSS 1018
976 TPKEBAQALDLDTFKELFOTRGHTEBAVAAGKTTKQPCSSPPES-----ADT 1024
1019 PRGKSRGAPPADKAPFAE-----AQKLRGDPPTWTSGLPFPVPPREV 1062
1025 PSTRRQPKTLEKRDVQKELSAKULTQTSGETTHTDKVPGBDKSINAF-----RET 1078
1063 IKASPHAPDPSAFVAPRGHP-----LPLGLHDTA-----RPVLRPPTISNPPPLISS 1111
1079 AK-----QKLDPRNASVYSGKRNKTKTEKAQPLDLGIMKELFOTPVCTDKPETHKTTIAC 1135
1112 AKHPSVLEIGAIISOGMSVOLHAYSE-----HAKAPVGVTVGLPLP--MDPKQLAP 1164
1136 RSQDPVDVPTPSKQSKSLRKVDVEEFPALRRKTPSAGKAMHTPRAVSGEKNIYAF 1195
1165 SGVQKQQLS-PRGAQAPRESICVPTAQSALRGTLALSVGGSITKQIBSTRPDSAI 1223
1196 MGTPOVKDLTENTLTGSKRRLQTPKE-----KAQALDLAG--FKELFOTRGHTE-- 1244
1224 TYRGSITHTPADVLVYKGTIRIIGEDPSRLDGRSDELPGHVIYEGKAGHVLVYSGG 1283
1245 -----SMTDKTAKAACSSQDPL--DKNPASSKRLKTSLSK--VGVEBELAVG-- 1291
1284 MSVTQCSKEDGRSSSGPHE-----TAAKRTYDMMEGVGRAISGASIEGLGRAIP 1336
1292 -KLQGTSETHTHTHEPTGDKSKMAFESPKQILDSAAISLTG--SKRQRTPKGKSEV 1347
1337 PER-----HSPHLKEQHINIGSTI--QGIRSYLEADEDYLIRAKULLKEGTPP 1386
1348 PEDLAGFIELFOTPSHTKE-----SMTNEKTTIVSYRASQPD-----LVDTPTSSK 1393
1387 PPSRDLTEAVKTAQALGPLKLKPAHEGLVATVKE--AGRSIHETPREELHNPDELPLAP 1443

A:Cross-references: GB:L01475; GB:L01476
A>Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A4
A:Cross-references: GDB:132673; OMIM:120131
A:Intron position: 2935-2937
A:Intron: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A>Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
A:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3
er trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxy
F:1-38/Domain: signal sequence #status predicted <Sig>
F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <Mat>
F:39-61/Domain: amino-terminal nonhelical, NH1 <NH1>
F:62-1466/Region: interrupted helical
F:94-96/Region: cell attachment (R-G-D) motif
F:145-147/Region: cell attachment (R-G-D) motif
F:189-191/Region: cell attachment (R-G-D) motif
F:310-312/Region: cell attachment (R-G-D) motif
F:724-726/Region: cell attachment (R-G-D) motif
F:785-787/Region: cell attachment (R-G-D) motif
F:989-991/Region: cell attachment (R-G-D) motif
F:1212-1214/Region: cell attachment (R-G-D) motif
F:1467-1690/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:1579-1666/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:47-52,55,57,266,400,460,492,494,666,750,828,1095,1131,1294,1317,1375,1407/Disulfide bo
F:142,669/Binding site: carbohydrate (Aan) (covalent) #status predicted
F:1480-1566,1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F:1525-1531,1634-1641/Disulfide bonds: #status predicted
F:1588-1663,1622-1666/Disulfide bonds: (or 1588-1666, 1622-1663) #status predicted

Query Match 2.6%; Score 339.5; DB 1; Length 1690;
Best Local Similarity 21.1%; Pred. No. 9e-05;
Matches 375; Conservative 107; Mismatches 617; Indels 681; Gaps 92;

761 DTGNGPK-PPATLGADP-----PPGP-----PTPRRTSRAPITPPASEATGATPP 809
380 ETDDVPPPPGILGPPGACAGMTGPPGPGILGPPGEGITGRDS-APGRGPR 438
810 PAPPSPAPPVVPKEKEETAAAPVEGEQKPPAAEELAVDTGK-AEEFVKSECTE 868
439 GSPGLPGAP-----GLQGPGSSVTCVGNPQPGIKGVAP 476
869 EAEGRPAKGDAAATAEAGLKAKEKGGSRATTAKSSGAPPOSDSATSADVD 928
477 PGGRGKGEKGNGLCACEPGRPPGPPGLPGRQSKDGLGPMVGTK----- 526
929 AEGGDGNRLSPRLITPTGDRANASPOKPLDKLKORAAIIP-----IOV 978
527 ---GD-----PGPPGAGPPGLP-----GKHGASGPPKMGKAGDMVVR 563
979 TKVNEPRDEADATKPAAPAPPONLQESDAPQPGSSPR---GKSRPAPADKEA 1034
564 VKGHKGER-----GPDGP-----GPPGPGSHGRDHAGEKDDPPPDHED 606
1035 FAEEAKLGD-----PCWTSGLPFPVPPREVIKASPHADPAFAVAPRGHP---L 1084
607 ATPGGKGFPPGLPFGKAGVGPGLGFPGRGE-RGNP-----GVGHHGVNG 654
1085 PLGIH-----DTAPVLPAPPTISNPPLISSAKPSVLEROIAGISQMSVOLHY 1135
655 PDLKQKQKGTISCNVTYPRGRHPGPDGP-----GPKGPPGGA--PLSGS--- 702
1136 PYSEHKAAPVGYVTMGLPLM-----DPKTLAPPSGVKQEQLSPPGQAGPPESIGVPT 1188
703 --DGHKGRPGTPTAETIPIGPBPGFRGDMGDP---GFGGEKGS--SPVGPPGPGSPGV-- 752

1169 AOEASVLRGTALGSV--PGSITTKGIPSTRYPSDSAITYRGSITHTGPADVLKGTITRI 1246
753 NGQKQIPEDPAFGHIGPPGKRGKLSGVPIKGP-----RGD--PGCP----- 791
1247 IGEDSPSRLL-----DGRGDSLL-----PKGHVIEGKKG-----HYLSYEGMSV 1286
792 -GAEPPAGIPGLGLKPKGRGHNAGFPVGPDPHSGERAGPILRGPGLPVGYG----- 846
1287 TQCKEDGRSSSGPHEHTAARITYDMEGRYGALISSASIEGLMGRAIPERRSPHHIK 1346
847 -----SPGAGGKGQF-----GVVGP-CPAGMKGLPGLGRPDHAPPGI- 886
1347 EGHNRGSITQGISPSYAEADYLRRAKLLKREGEPPPPSRDLTEAYVQALGPLK 1406
887 -----PGLPPGPD-----DGLPPGPPK-----GRG 909
1407 L-----KPAHEGLVATVKEAG-RSIIHEIPREIRHTPELPLARPLKEGSITQG 1454
910 LPGPPGPPGERKPKAEGCPGAKGEPGKMSGLPD-----RGLRGAKAIG 957
1455 TPLKYDTGASTGSKKHVRSILGSPGRTPPVHPDLVMADARALERACYESLSKSPGT 1514
958 PP--GDEGEMALISQK-----GTPGEPGPP-----GDDGPPGERGD 991
1515 ASSSGSIAARGAVIVPELKGPRQSPLYEDHGAPFAGHLPGSPVTRKEPTRLQEGSL 1574
992 KCTPPGQGRG-----ELGR-----YGP--GPHRGECEKQPPPPPPPP 1032
1575 SSSKASQDRKLTSTPREIAKSPHSTVPEHHPPISPEHLLRGVSGVDLYRSHPLADP 1634
1033 GSTGL--RGFIFP-----GLPDDQGPSPGPP--PGFGIDGARG----- 1069
1635 TSPPGIPLDAAAYLPRHLAPNTTYLHYPIYIRGYPTDPALENNQTIINDVITISQ 1694
1070 ---PKGNKQDPAS-----HFGPPG-PKEBPGSPGCG----- 1097
1695 MEHNATAMAGRADMLRGLSPRESSLALNYAAGPGIIDLQSVPLVLPPTPTPAT 1754
1098 -HFG--ASEEGILPGI-----GPPG-----SPRPPPPG-- 1124
1755 MDRLAVLPAPQPFSSRHSSSPLSPGPTHTLTKPTTSSSERDRDRDREREKSI 1814
1125 -----SSGP--PGCPDHPGILRGQPGEMD--PGPGLGDDPPI 1161
1815 LNSTTVEHAPLWRPCTEGSSSSSGSGG--GSSSRPASHAHQHSPIPRVDALQ 1872
1162 -----PPPGIKGPPSGSPGLNGHLKQKGTGASGLHD-VGP----- 1199
1873 QRPVLTHTGMKGIITAVPSPKTVLRSTSTSPVR-----AATPPATHCPAGTLDGV 1928
1200 --PGVIGIPLGK--ERQDPSPI-----SPGPPGKKPPPPGSSGPPG----- 1242
1929 YPTLMEPVLLPKPADRVAPERRPADTGHAFIAKTPRANSGLPASPSSKGSPPRLVPV 1988
1243 -----PAGATGRAPKDIIPDPGPDQG-----PPGPDGPGADGP-----PGL 1280
1989 SGHATIAATPAKILARNHASPDPAAPASDPHNEKQSKFSLQELTASLGYNG-- 2045
1281 PGSVDLIRGEPDC--GLPPPPGPPGPP-----GTXKPPG 1316
2046 -----SSYSPGEVE-FVSPVSSPSLTHDKGLRKHLELDKSHLEGELRKPQDPVYLGE 2099
1317 CQKQKQKPMKFPFGQGPNGFPGRGEGKGLG-----PPGRGPPGLPG- 1361
2100 AALHLPLRLPSQSSSP---LQAPGVKQKQVYTLAQHISEVITDYYTRHNPQLS 2156
1362 ---PRGEGPPADVDPCRIPLPGAPGRMRPBEAMGL-----PGKR 1400
2157 APLPAPLYFPGASC--PVLLRPPSDLYLPPDHGAPARGSPHSEGKSPPEPNKTS 2213
1401 GP-----PPGCGGEPGLDGRGVD--GVF--GSPPPPKPKGDT----- 1435

QY 862 V---KSECTEAEBSBPAGKXDAEAAEATAGALKAKEKGGSGRATTAKSSGAPODSSS 918
 DB 271 VRRRRSRLPKHSDDSGSKHDATASSSAAATVRA-----MSSVGIGTISCS 318
 QY 919 ATCSADEVDEABGCKNRLSPRPSELTPGTGPRANASPOKPLDLKOLKORAAAIPIQV 978
 DB 319 VQTEBDQ-----PRVSPA-----IHTNADDPKEI 345
 QY 979 TKVHEPPED-----AAPTKEPAPAP--PPQNLQPSDAPQOG-----SSPRKGRS 1025
 DB 346 VRYSAPFKTGRGESLACQTEPDQAGVAGFQVLGPTAISPYLEIGIQTVPGLGRREK 405
 QY 1026 PAPRDKAFAFAA-----AQKLPDPCWTSGLPFPFPPEVTKASPHADPSAFAP 1079
 DB 406 KKPPPLEIGYAHLPESLSQVSRQPKSPQVLISPSV-----LSHRLDLSFASSE 460
 QY 1080 ---RCHPLPLGLHDA-----RPVLPRP-----PTSNPPLISSAKHPSVLERQIGALS 1126
 DB 461 RLNRHNVBP-QKHFTADSLAQOTLPFRMKTLOKSLSPKELSPFAEE-SAKER----- 512
 QY 1127 QGMSVOLHVPYSEHAKAPVGPVTMGLPLPMDPKLAPSPGVKQEQLSPRGOAGPESLGV 1186
 DB 513 ---FSLYQH-----QGLGSGVSAALP-----PNSLV--RKVYRILPSP-----PREBAHL 552
 QY 1187 PTAQAS-VLAGTALGSPGSGITKGIPS--TRVEDSALITRGSLTHGTADVLKGTIT 1244
 DB 553 PLACGASQLYAASTL-----LQRLGTPPTVPATKA-----SLRELDRLD 593
 QY 1245 RIIGDPSRLDRGEDSLPKGHVIEGKKGHVLSBEGMSVYQCKSKEDGRSSSGPPHET 1304
 DB 594 RLVHEHS-TKL-RKQALDEBEKEIDAKLTL-----ELGITO----- 630
 QY 1305 AAPKRTYDMEGRVGRALSSASIEGLMGRALPPEHSPHMLKEQHHRGSIITGIPRSYV 1364
 DB 631 -----RKESLAKDRG--GRDYPPLGLGEH----- 653
 QY 1365 EAGEDYLK-REAKLLKREGTPPPPPSRDLTEAYTQALGPKLKPAHEGLVATYKEAGR 1423
 DB 654 ---RDYLSDELMQRLQSGCTTPAGQFVDF-----PATMAAPAT--PGGP 693
 QY 1424 SIHEIPRELBHTPELPLAPRPLKEGSIITQGTPLKYDTGASTGSKKHIVRSLISGPGRT 1483
 DB 694 TAFQOFRQ-----PPAP-----QYSAGS-----GGPQN 718
 QY 1484 FPPVPLDVMADARALERACYEESLSKSRPGTASGSGIARCAPVIVELGKPRGSPPLY 1543
 DB 719 GPPAH-----QAPYFPGSTYAPAPAPPGAS-YPAEPGLPNOQAFRP 759
 QY 1544 EDHGAPFAGHLPGRS-----PYTWREPTRLQBSGLSSSKASQDKLSTPRELAKS 1595
 DB 760 TGH---YAGQGTMPPTQGTLPFPVPDASRAPLOKPRQTL---ADLEKVPNTYEVIASP 812
 QY 1596 ---PHSTVREH---HHPHISPYEHLRGVGVDLVRSHIPLAFDPTSLPRGILDAAAY 1649
 DB 813 VVPMSSASSESYSPAVSSGIE---QGVV-----PEVPRAGDGSISQ----- 853
 QY 1650 YLPRHLAENPTYPHLYPYLLRGYDPTALENRQTIINDYITSQOMHNT--ATAMAORA 1707
 DB 854 ---SPAPYF-----SDSHYTSLEQVVPNNYVIMIDISELTKOSTSTADS 896
 QY 1708 DMLKGLSRES-----SLALNYAG-----PGIIDLQSVPLP 1741
 DB 897 QRLBPLGSGSGRPEKPEPGLDPLPCCYAAGESESESDSYDPHG-----KGGHJR 951
 QY 1742 VLVPEPTPTATAMDLAYLPTAQPFSSRHSSPLSCGPTHTLTKPTTSSSEERPD 1801
 DB 952 SM--ESNRPAST-----HYGDSVDRHGARVEKYPGEMGPHPKSLAPALSSKSKH 1005
 QY 1802 RERDRERERK-SILTSTTVEH--APWRPTEQSSSGSSSS----- 1841
 DB 1006 RKQGEQKISKFSFIEAKDVESDLASYPSPAVSSSLVSRGKRFODEITTYGLKKNVVEQ 1065

QY 1842 -----GGGGSSSRPASHS-----HAHQSPIS---PRTOALQORESVL 1878
 DB 1066 KYTGMSRDAVEDDRITGSSRSRAPSAAYGCEKLSHDFSGMGYERERAAVR-----L 1121
 QY 1879 HNTGMKGIITAVEBSK---PTVLNSTSTSSVPRPATTPPATHCPLGTL----- 1935
 DB 1122 QKAGKPSLSLMAHSRVRPFRMSQASEESVSLGRPRA-----GGPLRPGDTCQF 1176
 QY 1926 --DGVYPTLMEPVLLPREKAPRV---ARPERRADTGIAFLAKPRA-RSGLEPASPCKGS 1979
 DB 1177 CSHSMBDVOEHV---KDGPRNAHYKKEEGYIILDSHCVSDSAHYHLGQEBTMDPKPR 1233
 QY 1980 EPR-PLVPRVSGHATARTAPKXLAHPHNASDPRAPASADPRREKTQSKPFIQETEL 2038
 DB 1234 DARSDRRNHGNA-VSSSGCKGRPAHSTHYDVEPEBGLMWRDEGGRGHNASAKHRH 1292
 QY 2039 RSLGHSYSPEGEVEVSPVSPSLTHDKLPRLHELDKSHLEGELRPK--OPG----- 2092
 DB 1293 GDHRHSGRHT--GEFGRAAPK---HARDLGHARPHSQPSAPAMPKKGQGVYPS 1347
 QY 2093 -----PVKLGEMAILPHLRPLPESQPSPL---QTAPGVKGNORV 2132
 DB 1348 AEYQPSRASSAYHNASDSKSGRQAHSGRALQSKAEPOKOPOLQGRQAPG----- 1400
 QY 2133 VTLAQHISEVITQDYTHHPQQLSAPLPAPLYSPPGASCPVLDLRPPSDLYLPPPH-- 2190
 DB 1401 -----PQOSQPSRQTPS--GAASROPQTQOQOGLQPPQOAL 1439
 QY 2191 -----GAPARSPHSEGGKRSPEBNKTSVLGGEDGIEVPSPREGMTEPG 2235
 DB 1440 TQARLQOOSQPTTGSAPASQF--AGKQPGHSTMT-----GQPAPFPABEQNG 1489
 QY 2236 HSRSAVYPLLYRDEQTEPS-----RMGSKSPGNTSQPPA-----PFSKLTES 2278
 DB 1490 SKGTAKAPQOGR-APQAPPTPGPAGVAKGAR-PGGTPGAPGADGSESVSKILPG 1547
 QY 2279 NSAMVSKKOE 2289
 DB 1548 GAEOAGKLTB 1558

RESULT 69
 CGHUB
 collagen alpha 4(IV) chain precursor - human
 N.Alternate names: procollagen alpha 4(IV) chain
 C.Species: Homo sapiens (man)
 C.Date: 06-Feb-1995 #sequence revision 03-Oct-1995 #text_change 16-Jun-2000
 C.Accession: A55360; S36854; S28777
 R.Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reeder, S.T.
 J. Biol. Chem. 269, 26172-26177, 1994
 A.Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp
 A.Reference number: A55360; MUID:95014445; PMID:7523402
 A.Accession: A55360
 A>Status: nucleic acid sequence not shown
 A.Molecule type: mRNA
 A.Residues: 1-1690 <LEI>
 A.Cross-references: GB:X81053; NID:G574805; PIDN:CAA56943.1; PID:G574806
 R.Sugimoto, M.; Ohashi, T.; Yoshioke, H.; Matsuo, N.; Ninomiya, Y.
 FEBS Lett. 330, 122-128, 1993
 A.Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha
 A.Reference number: S36854; MUID:93374047; PMID:8365481
 A.Accession: S36854
 A.Molecule type: DNA; mRNA
 A.Residues: 1219-1658, 'FE', 1661-1690 <STG>
 A.Cross-references: DDBJ:D17391; NID:9440365; PIDN:BA04214.1; PID:9457161
 R.Experimental source: whole eye
 R.Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
 J. Biol. Chem. 267, 23753-23758, 1992
 A.Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
 A.Reference number: S28777; MUID:93054733; PMID:1429714
 A.Accession: S28777
 A.Molecule type: DNA
 A.Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>

QY SEQEEETRE-----KFMQHPKNGELINSF-----LERK 461
 Db LELETAARFKLLEGEESRIELVHPMTTGTREAVPELITNGKASADDDSSKOGTVRAK 490
 QY 462 TVAECCVLYYYLTKKNENYKSLVRASVRRRGKQ00000000000000000000PMBRSSQEEK- 520
 Db 491 SVSPRVV-----AETKLTSTSYCGGDEADDEGKSDSTHTHAEAEFRADSDAT 542
 QY 521 ----DEKEKEAEKEEKEPEVENDKEDLLKEKTDTDTSGEDNDEKEAVASKRKTANSOG 576
 Db 543 GTGLDEVKEESVLKEEKKQSVKRDDEEEEDNDQTESE-VDPEADADIEKQKESD-- 599
 QY 577 RRRKGTITSMANBANSEAHITPOQSAELASMLINSSKRTTEEMETAKGLLEHGRNWA 636
 Db 600 --KTESDRAESRRASVRAITPESBDKAEVSYSKSS-----ITSEAFIDKS----- 645
 QY 637 IARWVGSKTVSQCKNFYFNKKRKQNLDELIOQHKL---MEKENARAKKKKAPAAASE 692
 Db 646 -PKSGSGSR-----DTLFQGDKTYTPVLLRANSLQCAPTEKP--LSD 685
 QY 693 EAAPRVVEDEEMEASGV-SGNEEBEWEAEALHAGNEVPREGCGSPATVNNSSDTESI 751
 Db 666 TIKSPVSEPAISPVSVSGAETSPAERMTSPISGSG-----SKASPVASEAT 736
 QY 752 PSPHTEAKDQYQNGPKPPATITGADGPPGP---PTPRKTRAPATETPTASEATGAT 807
 Db 737 KSPVSE--KSASGSKSPVPSERALSPLSENSVSHSTANSHTRG--PTASEKSVKSPJT 791
 QY 808 PPPAPSPSAPPVVPVPEKEEBETAAPVERGESEBOKPPAAETLAVDTGAKEPVSECT 867
 Db 792 HSERLASFPAKSPITMEPAKS-----PKDSEKELSPERSE-SVMSG-----KSQIT 837
 QY 868 EEAEBCPAKGGKDAEAEATAGALAEKKEGGSG-----RATTAKSGAPODSDSA 919
 Db 838 SSSAKSPVPSEKAD--SEKSATSPTPSEKVDSSAKSPTHSDRSQSQKARSPTMSDHIX 896
 QY 920 TCSADEVDEAGGDKRRLSPRPSLITPGDPRANASPOKPLDKOLKORAAALPIPTQVT 979
 Db 897 SPIDEEKSESEKALSYSKSDHDKSPVPEKEASEKARSAPVPESEKASEKARSQVISE 956
 QY 960 KYHEP--PREDADPTPAPAPAPPONLQPESDA---PQPPSSPPGKRSRAPPADKXA 1034
 Db 957 PAKSPVPSEKAESEKARSAPVPESEKASEKARSFVPEKEASEK-SARSVP-PSEKO- 1013
 QY 1035 FAABEAOQLPGDDPCWTSGLPFPVPRAEVKA-----SPH--APPSPAFSVYAPGHPPL 1086
 Db 1014 ---BEKKAKSTSTSQARSPVPESEKASEKARSAPVTSIHVKSVPVSEKSEBEKARSQV 1070
 QY 1087 GLHDTARPV-----LPRPPTISNPPLISSAKHPSVLERQIGALISQMSVOLHVPY 1137
 Db 1071 TSEHYKSPVPESEKASEKAKSPVASEP-----AKSPFPESEKASEKAKSPVT----- 1119
 QY 1138 SEHAAPVCPVTMGDPLPMDPKKLAPFGSGVKOQOLSPPQDAPPBESLQVPTQOASVLRG 1197
 Db 1120 SEHAATSPVAS-----EKSESGSKSPAASEKASEKSKSP-----VPSEKESKTAG 1166
 QY 1198 TALGSPVGGSLTKGIPSTRVPS 1219
 Db 1167 SP-----GSSDEGILTTTTS 1182

RESULT 67
 S21626
 collagen alpha 1(I) chain precursor - mouse
 C.Species: Mus musculus (house mouse)
 C.Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #ext_change 13-Aug-1999
 C.Accession: S57243; S16374; A23982; 149559; 149557; S39789; 148300; S21626
 R.Li, S.W.; Khillan, J.; Prockop, D.J.
 Matrix Biol. 14, 593-595, 1994
 A.Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
 A.Reference number: S57243
 A.Accession: S57243
 A.Molecule type: mRNA

A:Residues: 1-1453 <LIS>
A:Cross-references: EMBL:U08020; NID:g470673; PIDD:AAA88912.1; PID:g470674
R:Meisner, M.; Tomam, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:g50484; PIDD:CAA1046.1; PID:g50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A>Title: Nucleotide sequence of a cDNA clone for mouse procollagen(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:Cross-references: GB:M14423; NID:g192261; PIDD:AAA37333.1; PID:g192262
R:Monson, J.M.; Friedman, U.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: 149559; MUID:83141374; PMID:6298587
A:Accession: 149559
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:g192263; PIDD:AAA37334.1; PID:g192264
R:Habbers, K.; Kuehn, R.A.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A>Title: Insertion of retrovirus into the first intron of alpha(I) collagen gene leads
A:Reference number: 149557; MUID:84170331; PMID:6324198
A:Accession: 149557
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-25 <RE2>
A:Cross-references: GB:K01688; NID:g192246; PIDD:AAA37330.1; PID:g553881
R:Penton, S.P.; Lamande, S.R.; Hannagan, M.; Steacy, A.; Jaenisch, R.; Bateman, J.F.
Biophys. Acta 1216, 469-474, 1993
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80,'E',82-105,'D',107-185,1031-1201,'G',1203-1218,'E',1220-1221,'T',1223-1231
R:Phodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A>Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-80,'E',82-105,'D',107-147 <REF>
A:Cross-references: EMBL:X54876; NID:g50486; PIDD:CAJ8657.1; PID:g50487
A:Genetic:
A:Gene: COL1A1
A:introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
P:1-32/Domains: signal sequence #status predicted <SIG>
F:23-151/Domains: amino-terminal propeptide #status predicted <PRO>
F:30-89/Domains: von Willibrand factor type C repeat homology <VWC>
F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F:1224-1453/Domains: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.6% Score 341; DB 2; Length 1453;
Best Local Similarity 21.7%; Pred. No. 6.7e-05;
Matches 335; Conservative 89; Mismatches 539; Indels 580; Gaps 80;

707 ASGVSGNHEEWEEBAALHAGNEVPAGE-----C-SGPATVNNSSTTESIPSP 754
DB 15 ATALTTHGQDEIPEVSCGH-NGLRVNGERTMKKEPVCLICTCHNCFTAVCDVDQCNEBIDCF 73
755 HTFA-----AKDTGONGPK-----PATLGADGPP--PGP 782

```

QY 1126 SOGMSVOLHVPSSEHAKAPVGVITWGLPLMDPK-----KLAFFSGVK-QEOL 1172
    |||
    |||
    |||
Db 3013 PKLQSVQ-----EKGSVPRVVK-----PLDPKPOKEQESKNEKNEKVPSTNVLDDQA 3060
    |||
    |||
    |||
QY 1173 SPRGAGPESIGVPTAOEASVLRG-----TALGSPGSGITKGIESTVSPDSALTYGS 1228
    |||
    |||
    |||
Db 3061 SSSSNQGGSGSAKPSASIFTLPRKKEKKMTSTAESSTIPITSPENQEAETSTNKS 3120
    |||
    |||
    |||
QY 1229 ITHGPADLVYKGTTRITIGEDSPSRDLRGREDSPKGVHYEGKKGHVLSYEGMSVTO 1288
    |||
    |||
    |||
Db 3121 -----PIERIKKPTGOLIKRSPSDVPRKCV-----SGALVPRVDKPI-S 3163
    |||
    |||
    |||
QY 1289 CSKEDRSSGPPHETPAKRTYDMWEGVGRASISASIEGLMGRALIPERHSPHLLKQ 1348
    |||
    |||
    |||
Db 3164 -----NSPSYKPMARRPKA-----DSAADSGDLG----- 3190
    |||
    |||
    |||
QY 1349 HHIRGSITGCIPTASVYAEODVIRRAKLLKREGTPPPPPSRDLTAAYTOALGPLK 1408
    |||
    |||
    |||
Db 3191 -----SDEGSTSHGCLKDENSQKP-----AKTIVRFGFISTE 3225
    |||
    |||
    |||
QY 1409 PAHEGLVATVKEAGRSIHIEPRELHTPEL-----PLAPRPLKE-----GSITQOTPLK 1458
    |||
    |||
    |||
Db 3226 PPPAPAPQV-----ELPKVE-----PTPEISFESLAQQAOKTLAKEASKILEHYPT 3275
    |||
    |||
    |||
QY 1459 YDTGAST-----TGSKKHVRSLLIG-SPGRTPPVHPPLVMDADARA-----LERACTEE 1506
    |||
    |||
    |||
Db 3276 EOTSPTAEVSPETKSGK-----LSFFGLSP-----FGKVTPKNTAQTOSTLVLESTSP 3328
    |||
    |||
    |||
QY 1507 SL-KSRPGTASSSGGSA-----RGAPVIVPELQPRQ-----SP 1540
    |||
    |||
    |||
Db 3329 LTPSPNETSNKSTSSKOPKSLSEKTIATOMIGTRGP-----IRKLSHRQKIGDPP 3386
    |||
    |||
    |||
QY 1541 LTTEHDGAPAGHLP-R-GSPVTKREPTPRLOESLSSKASQORULTSTPREIANKSPST 1599
    |||
    |||
    |||
Db 3387 LQJSGDS-----LPRAPGSPKPPQLAEVQGVQTVNPPSQKRYTRLGREKKPPPKAK 3441
    |||
    |||
    |||
QY 1600 VPEHHHPPISPYEHNLKGVSGVLYKSHIPLADPPISIPGILDAAYLYP----- 1652
    |||
    |||
    |||
Db 3442 TPLSIRVPOS-----GVS-----ISTP-AIDLSAKEVILADIEKTKM 3478
    |||
    |||
    |||
QY 1653 RHLAPNTVPLVYPVILRGVPTAL-ENROTIINDYITSOQ----- 1694
    |||
    |||
    |||
Db 3479 RKRKPKMETTL-----VLDGTEHFKULTAENLQLSQPKSKKKKKKTKTWISVIGCB 3534
    |||
    |||
    |||
QY 1695 -----MHNTATAMAQRADMLRGISPRESSIALNTAACPRIIDLSQVPHLPVLP 1747
    |||
    |||
    |||
Db 3535 NDSITKCMQHN-----KLLEFRDLKKSILITRQ-----LDEKMI----- 3568
    |||
    |||
    |||
QY 1748 PGRTATAMDRLAYLPTAPORPSSSHSSPLSPGGPTH/LKPTTSSSERERDRDR-DR 1806
    |||
    |||
    |||
Db 3569 -----TAPAVFNPRF-----TBSSTIKHLLANKPKFEF 3597
    |||
    |||
    |||
QY 1807 DREKESILNSTTVEHAPITWRPCTEOSGSGSGSGSGSSSRPASHAHQHSPISPR 1866
    |||
    |||
    |||
Db 3598 TRKKLKLANN-----RQCKIRMSG-----HNSQPKDVAISAI 3631
    |||
    |||
    |||
QY 1867 TODAQQPVSILANTGKGIITAVEPKPYLRTSTSSPYRAPATPPATHCPLOGTLD 1926
    |||
    |||
    |||
Db 3632 RQ-----VOKVFREFOSTGKKMNPILISGFR-----KSLIKPPKPRATV-----LFRIDSEGN 3679
    |||
    |||
    |||
QY 1927 GVPYTLMEPVLLPEAPRVARPER-----PRADTGHAFI-----AKPARAGLEPASPSKSGE 1980
    |||
    |||
    |||
Db 3680 KLINIKLGPDMQKNDKCNPMAMBRKKSNSYLKRNAMAKWAKYG-----KKNMTLFE 3736
    |||
    |||
    |||
QY 1981 PRPIVPPVSGHATTARPPAKNLAHHASDPDP-----APPASADPHREKTSQKPSIOELE 2037
    |||
    |||
    |||
Db 3737 TNOJLRRLG-----APKDSHPDVLVDPKADAPPSPRS-----LSWTTVSSNANSL- 3786
    |||
    |||
    |||
QY 2038 LRSIGYHSSSYSPBEVPEVSPVSSSLTHDKGLPKHLELDKSHLEGLRP-----KOP 2091
    |||
    |||
    |||
Db 3787 LDALNTVGS-----EPSTSSGIVTDIAPIVKEQFRKAK-----ESRPTQOTVSKY 3834
    |||
    |||
    |||

```

```

QY 2092 GPKVAGEEAHLPHLRLEPSSQSSPLLQTADEKVGQGVVTLAQHISEVI----- 2143
    |||
    |||
    |||
Db 3835 PTVYLGNTVVR-----SSSSSTSL-----VNLDEREAVSSLAQDSGSIYEVGQEEB 3881
    |||
    |||
    |||
QY 2144 TQDYTRHHPOQLAPLAPLYSPGASCPV-----LDLRPPSDLYLPPDHGAPARQSP-H 2199
    |||
    |||
    |||
Db 3882 VETTKRFPDPFVAPISALFKAFPTSSPEVGLBADRSRPTERTL-----RGLFRSSPTM 3937
    |||
    |||
    |||
QY 2200 SEGGR-----SEPKNTSVLGGGEGDIEVSPPEGTEGHSRSAVYPLVR-----DGEQTEPS 2255
    |||
    |||
    |||
Db 3938 PEFSTRVSSSPTRRRNRG-----PVTSVQSSAEDILALMOAQPDALENMEE 3987
    |||
    |||
    |||
QY 2256 RMGSKSPGNTSQPPAFPSK-----LTESNANVYKSKQOINKKLTNNRNEP 2302
    |||
    |||
    |||
Db 3988 QVSKFPPIQALAVLNGKGMWRPKLKNLSLYQFKELEKESKRXLYEVLQNYKNNK 4047
    |||
    |||
    |||
QY 2303 EYNISQPTIELFNPAITGTLMTYRSQAOVHAETMGLAIIKRLMKRYQWESSPP 2362
    |||
    |||
    |||
Db 4048 SYSTLQI-----LITFN-----LDQYKRTNLYRPPKLTNDLGLKIDP----- 4084
    |||
    |||
    |||
QY 2363 LSANAFNPLNAS 2376
    |||
    |||
    |||
Db 4085 -----NSLRVSS 4092
    |||
    |||
    |||

```

RESULT 66

A:6194

neurofilament protein NF-220, high-molecular-weight splice form - longfin squid

C:Species: Loligo pealeii (longfin squid)

C:Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Mar-2000

C:Accession: A6194

R:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Garner, H.; Batty, J. Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992

A:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod d

A:Reference number: A6194; PMID:9235751; PMID:1379729

A:Accession: A6194

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:molecule type: mRNA

A:Residues: 1-1200 <MAV>

A:Cross-references: GB:M94389; NID:g161291; PID:g161292

A:Experimental source: stellate ganglion

A>Note: sequence extracted from NCBI backbone (NCBI:113499)

C:Superfamily: neurofilament triplet H protein

C:Keywords: alternative splicing

Query Match 2.6%; Score 341; DB 2; Length 1200;

Best Local Similarity 19.4%; Pred. NO. 5.5e-05;

Matches 237; Conservative 196; Mismatches 491; Indels 298; Gaps 47;

```

QY 137 KORSITGKLEVPSPSPHTDELVPPRLSKELIQMDRVRE-ITM---VEQGISK 192
    |||
    |||
    |||
Db 120 QNKKLAGELEBKSKWKETSAIKEMYTEL--EARKLIDATNKKITLIDRVTELIDQ 177
    |||
    |||
    |||
QY 193 LKKKQOQLEEEAAKPEPEKPVSPPIESKHSIVQIYDENRKAFAAHLILEGIAPQV 252
    |||
    |||
    |||
Db 178 LERQOQDLEES-----RTYQIDQEQIARQNOQLAD-----LEQ----- 211
    |||
    |||
    |||
QY 253 ELPLYNQPSDT-----ROYHENIKINQAMRKLI-----LYF 284
    |||
    |||
    |||
Db 212 EISMRLRSIESLEKEMQSNITLAKONDEMQRMDLNNETINHLDAENRQTLDEEELF 271
    |||
    |||
    |||
QY 285 KRANHA-----RKQKQKCCQYDQMLELLEKKVRIENPPRRAKE 326
    |||
    |||
    |||
Db 272 QKQVHAQELKEALALAYRDTTAENEFWNEIQAIRDIQOEYDAKCDQMRD----- 324
    |||
    |||
    |||
QY 327 SKVREYEE-----KOPPEIRKQELQERMOS-----RVGQSGSLSNASASEHEV 372
    |||
    |||
    |||
Db 325 --TEAYYNLKVQEFRTGATKQMEVYRNKEBNTYKLSMNTETIRNLADLEAANAQIERTN 382
    |||
    |||
    |||
QY 373 SEIIDGLSEQENLEKQROLAVIPMLYDADQRIKFINMNGIADPMKVVYDROVMNMV 432
    |||
    |||
    |||
Db 383 QDLRLDLEKDRQNE-----LESCQYKEEITKLRGEMESILKEIDLDMDIKLS 430
    |||
    |||
    |||

```

[illegible][illegible]


```

Db      376 HGAQGP---PQPPGNGSPG-----GKGMGPAGIPGAPGLMGARGPPGP 418
Qy      1148 V-TWGLP-----LPMDPKLAPSGYKQOLSPGQAGPPESIGVPTAGASVLKGT---- 1198
Db      449 AGANAPGRLRGAGGEPGK-----NGAKGEP-GPRGEGGAGIPGVPAKAGDEKDGSPGEP 473
Qy      1199 AIGSPGSGITGPISTRVPSDAIYRSGISITHGPRADVLVYKGTITRIIGEDSPRLSD 1258
Db      474 GANGLPGAAGERGAPGFRGPAG-----PVGILP-----GEGKPGAG-ERG 510
Qy      1259 REDSL-PKGVHYEGKKGHVLSEGMSTYQCSKEDGRSS-----GPRHETAAPKATYDM 1313
Db      511 APGAPGPRGAGBPGDQ-----VPGGPRKMGMPGSGPGSGDKKPPPS----- 556
Qy      1314 MEGVGR-----AISASIEGLMGRAIIPPERH-SPHILKEQHINRISTIGIPRSTVEAQ 1367
Db      557 -QGSGRPPGPPGSGRQPGVWGFPGKNDGAPKNGERGGPGGPGQGGPKNGETG 615
Qy      1368 EDYLREAKTLKREGPPPPPPSRDLTEAYKTQALGRLTKRAHEGLVATVEAGRSIH 1427
Db      616 P-----QGPPTGPGDKGDT-----GP-----PQPGLQGLPTGCP----- 649
Qy      1428 IPREILRHTPELPLAPRLKESITQGTLYKDTGASTTGSKKHVRSILGSPGRTFPPV 1487
Db      650 -PGENCK-----PGEBCP-----KGDAGAPGAPGKGDA-----GAPGERGP- 686
Qy      1488 HPLDWMADRALERACYESLSKRPCTASSSGSLAKGAPVIVPELCKRQSPITYEDBG 1547
Db      687 -----GLAGAPGL-----RGA--GPP--GPEGKGAAGP-----PG 714
Qy      1548 APFAGHLPRGSPVYMEPTPLQEGSLSSKASQDKLSTPRLIAKSPHSTVPEHHP 1607
Db      715 PEGAA-----GTRELQMPGER--GGLSGPGRGDKGEGEGP----- 750
Qy      1608 ISEYELHMGVSGVDLYRSHIPLAFPTSI PRGIPDAAAAYVLPRLHLPNTYPLYP 1667
Db      751 -----ADGVPGKD-----GPRG-----PTGP-IGPP 770
Qy      1668 YLIRGYPDTAALENRTIINDYITSQOMHNTATAMAQADMLKGISPRESSIALNYAAG 1727
Db      771 -----GPAQGRQDKKEGAP-----GLPGIAG 792
Qy      1728 PRGIIDSLQVPHLVLPPTPGTPA--TAMDRLAYLTPA-----QPRSSHSSSP--LSG 1780
Db      793 PFG-----SPGERGTGPPGAPFGGAGQNGRPGCKERKAPGKRG 836
Qy      1781 GPTHLTKPTTSSSERERDRDRDRERREKSLITSTTVHAPIMRGTBOSSGSSGS 1840
Db      837 GPPGVAGPPGGS-----GPAQPPGPGQVKGERS 865
Qy      1841 SGGGG-----GSSSRPASHSHAHQSPISPTODALQORPSVLANTGMGIITAVPSKP 1895
Db      866 PGQPGAGPFGARGLGPPGSGNGPGRPGSGSGKDGPPGAGNGAPG----- 915
Qy      1896 TYLRSTSTSPV-----RPAATFPATHCPLIGLIDGVYPTLMEPVLLPKAPRY 1945
Db      916 ----SFGVGCPKCDAGQPEKSGPAGQFPAGPRLG--IAGI----- 952
Qy      1946 APERERADTGAFLAKPPARSGLEBPASSPS-----KGSERPLVPPVSGHATTARTAPKN 2001
Db      953 -----TGARGLAGPPGMPG--PRGSPGQVKGESGKP-----GANG 987
Qy      2002 LAPHNASPPR---PAPPASADPHREKTOSKPSIOLELRSIGHYGSSYSPEGV---E 2054
Db      988 LSEGERPPGPQGLPGLAGTAGBERGDNPG-----SDGLRGDGSFGKDKRGE 1036
Qy      2055 PVSVPSPSLTHDKGLPKHLELDKSHLEGLERPKQPGVPLGSEAAHLPHLRPLESOP 2114
Db      1037 NGSP--GAPRAGHPGPPGVPGAPKSGCDGSESP--AGPAGAPRAGS-----RGAQPG 1089
Qy      2115 SSSPLLQT---APGVGHQRVVTLAQLHISEVITDQYTHHPQOLSAP-LPAP-----LY 2164
Db      1090 PRDCKETGERGAAGIKGH-----RGFPGNPAGPSFGAGQOGAI 1130

```

```

Qy      2165 SPPGASCPVLDLRP-----PSDLYLPPPDHGAPARGSPHSGKRSPP 2209
Db      1111 GSPGAPG---RGVBSGPPGKDGTSGLPDPGPPGRNGRSGSGSGHPGQPGP 1186
Qy      2210 -----NKTSLVG-GGED-----GTEP-----VSPPEGTEPGHSRASY 2242
Db      1187 PGPAPGPPCCGVGAALAGIGERKAGGAPAYYDEPHMDFKINDELITSLKSNQGLE 1246
Qy      2243 PLVYRDGEQTEPSR 2256
Db      1247 SLISPDGSRKNPAR 1260

```

RESULT 64

```

OFHGH
Neurofilament triplet H protein - human
N:Alternate names: neurofilament protein, 112K
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
C:Accession: S00979
R:Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
EMBO J. 7, 1947-1955, 1988
A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
A:Reference number: S00979; MUID:88328981; PMID:3138108
A:Accession: S00979
A:Molecule type: DNA
A:Residues: 1-1020 <LEB>
A:Cross-References: EMBL:X15306; NID:g35028; PID:CAA3366.1; PID:g1841430
A:Note: It is uncertain whether Met-1 or Met-2 is the initiator
C:Genetics:
A:Gene: GDB:NEFH
A:Cross-References: GDB:120225; OMIM:162230
A:Map position: 22q12.1-22q13.1
A:Introns: 295/1; 361/3; 403/2
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
F:1-100/Domain: amino-terminal <NTD>
F:101-410/Domain: rod #status predicted <ROD>
F:411-1020/Domain: carboxyl-terminal <CTD>
F:502-826/Region: 14-residue repeats
F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
(covalent) #status predicted
F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

```

Query Match 2.6%; Score 342.5; DB 1; Length 1020;

Best Local Similarity 20.6%; Pred. No. 4,1e-05;

Matches 201; Conservative 147; Mismatches 398; Indels 231; Gaps 32;

```

Qy      166 RLSKEELIONMDRVREIMVQQLSKKKQOOLEEBAKPPPEKPVSPPIESKRS 225
Db      168 RLQEHLLDIAHV-----RQRDDA----- 189
Qy      226 LVQIYDENRKAQAARHL-----EGIGPQVEL-----PLYNPSDTRQYH----- 268
Db      190 -----RQREBAEAARALARFAQEAARVLDLQKAQALQEGCYLRHHQEEVGE 242
Qy      269 -----NIKINQMRKKLILYFRKRNARK--OMKQKFCQRYQ 304
Db      243 GOIQSGAAQOMQATRDALKCDVTSALR---TRAQEGHAVOSTLQSEEMFRRLLR 299
Qy      305 LMEALEKVERIENPPRRAKESKYREYEQPEIRKRELOERNQSHVGQSGLSGS 364
Db      300 LSEBAK-----VNTDAMKSAQETLEY-----RQIQATTELEALKSTYDSJE 343
Qy      365 AARSEHVESEIIDGSEQENLEKQKROLAVIPMLYDADQRIKFINMGMLAADPKYK 424
Db      344 RQRESELDHQADIASYQEAIOO-----LDDELRTWKV-----EWAADLRBYO 386
Qy      425 DRQVMMMSQEKETPREKFMHPK---NFGILASFLKETYAECLVLYYLTKKNENYKS 481
Db      387 DLINVMALDITIAVR-KLLEGECRIGFGPIPSLPGRLKIPSVSTHIVKSEBKIK 445

```

A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A:Reference number: 155349; MUID:91161621; PMID:1672129
A:Accession: 155349
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 537-605 <EEx>
A:Cross-references: GB:M59312; NID:G180815; PIDN:AA52041.1; PID:G180816
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from cy
A:Reference number: A90438; MUID:80198282; PMID:6246925
A:Accession: A90438
A:Molecule type: protein
A:Residues: 728-895,'A',897-964 <SEV4>
A:Experimental source: liver
R:Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A:Reference number: A38303; MUID:91009133; PMID:2145268
A:Accession: A38303
A:Molecule type: mRNA
A:Residues: 861-1015 <COL>
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA59383.1; PID:G
R:Marko, B.S.; Dalgleish, R.
Nucleic Acids Res. 16, 2337, 1988
A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A:Reference number: 802119; MUID:88189827; PMID:3357782
A:Accession: 802119
A>Status: translation not shown
A:Molecule type: mRNA
A:Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>
A:Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA2986.1; PID:G30054
R:Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from cy
A:Reference number: A90446; MUID:81208139; PMID:7016180
A:Accession: A90446
A:Molecule type: protein
A:Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',1157-
A:Experimental source: liver
R:Lojdl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
A:Reference number: A93551; MUID:85087944; PMID:6096827
A:Accession: A93551
A:Molecule type: mRNA
A:Residues: 1065-1155,'P',1157-1466 <LOT>
A:Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1
R:Miikulin, M.; Dalgleish, R.; Kuive-Beckerman, B.; Remard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A>Title: Human type III collagen gene expression is coordinately modulated with the type
A:Reference number: 152393; MUID:86187804; PMID:3754462
A:Accession: 152393
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1161-1200 <MIS>
A:Cross-references: GB:M13146; NID:G180415; PIDN:AA52003.1; PID:G180416
R:Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A:Reference number: 159025; MUID:85216505; PMID:3858826
A:Accession: 179359
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1165-1196 <EMA>
A:Cross-references: GB:M1134; NID:G180417; PIDN:AA52004.1; PID:G180418
R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sjöpol, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha (III) collagen. F
A:Reference number: A92516; MUID:85157600; PMID:2579949
A:Accession: A92516
A:Molecule type: DNA

A:Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>
A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
A:Experimental source: liver
A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
ation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O
C:Genetics:
A:Gene: GDB:COL3A1
A:Cross-references: GDB:118729; OMIM:120180
A:Map position: 2q31-2q31
A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine r
C:Function:
A:Description: structural component of extracellular fibrous polymer that maintains inte
A:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coll; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>
F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:154-167/Region: amino-terminal nonhelical telopeptide
F:168-1196/Region: helical
F:1091-1093/Region: cell attachment (R-G-D) motif
F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi
F:161-1212/Modified site: allysine (Lys) #status predicted
F:263-284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F:163/Binding site: carboxylate (Lys) (covalent) #status experimental
F:584-1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted
Query Match 2.6%; Score 343; DB 1; Length 1466;
Best Local Similarity 21.1%; Pred. No. 5.8e-05;
Matches 357; Conservative 90; Mismatches 599; Indels 648; Gaps 82;
QY 715 ERMVEEALMAAGNEVVRGEGSPATVNSDTEISPTTEAKOTGONGPK-----PP 770
DB 63 DDIICDDDELDCNPDEIFPGCCAVC-----PQPTAPTRPNNGGPGCGKDDPDP 114
QY 771 ATLGAAGPP-----GPPTPPRTSRAPLEP---TPASBA-----TGAPT 807
DB 115 GIPGNNGPQIPQCGSPGSGPGPCIGSCFTGPNYSPQDSDYKSGVAVGLAGYFG 174
QY 808 P-----PPAPP-----SPAPPPVVRKEKEEETAAAPVEEEOQPPAAEELAVDT 855
DB 175 PAGPPGPPGPPGPGTSGHPSPGPGYQPPGEGQGPFGPP-----GPPGA----- 220
QY 856 GKAEPPVNSECTEAEAGPAGKDAEAAEATAGALKAEKKGSGRATTAKSGAPQDS 915
DB 221 -----TGPSSGA-GKDS-----SGRP----- 236
QY 916 DSSATCSADEVDNAGCGDKNRLLSPRLILPTGDP-ANNSPQKPLDKQLKQRAAIP 974
DB 237 -----GRGERGLPDPFGIKGPAGIGPFGMGKCHRGDFGRNGEKETGAP 281
QY 975 PIQVTKHPPPPEDAAPIKAPAPAPPQNPQNPESDAQQPSS-----PRGSRSPA 1027
DB 282 GLK-----GENGLPENGAPGPMGPRGAPGEGRGRLPAAAGANGDARGSDGPG 333
QY 1028 PPADKEAPFAAEQKLPDPPCWTSGLPVPPVPRVYIKASPHAPDASAFYAPPGHPLPG 1087
DB 334 PPG-----PEETAG--PPGPGAGGVGPAGSPSSNAGAPGGRGPPGQ 375
QY 1088 LHDTPAPVLPPEPTISNPPLISAKHPSVLERQIGAISQGMVOLHVPYSEHAYAPVGP 1147

```

QY 1719 SLALVYAGPRGIIIDLSQVPHLVLPPTGTATAMDLATVLPAPCPSSRHSSPLS 1778
      |||||
      ---GLGIAGPRG-----GPG-----EGEGHPREPAPF----- 810
Db 785
QY 1779 PGGPHLTKPTTSSERERDRDREREKSLITSTTVEHAPWPRPTEOSSGSS 1838
      |||||
      ---CAKGRGAPGEGGPGPPGPGTSS-----GPAPPGPGQVKGGR 862
Db 811 PGAPQNGP
QY 1839 GSSG-----GGGSSSRPASHAHGHSPISTPTQALQORPSVLNTMKGITAV 1890
      |||||
      GSPGPPGTAAGPPGGGGLPGRPANNNGP---GPPGSGAPKDGKPPGPNNGS----- 911
Db 863
QY 1891 ESKPTVLRTSTSTSPV-----RPAATFPATHCPLGTIDGVYPTLMEVLLPKAPRVA 1946
      |||||
      912 SPGNGLIAGPKDADQPGKCPGAGCPGSGPGLG--IAGL----- 951
Db 1947 RPERPADTGAHFLAKPPARSGLEPASSPS---KSEPRPLVPVSGHATTARTPAKXL 2002
      |||||
      952 -----TGARGLAGPPGMPG--PRGSPGQIKGESGKP-----GA 984
QY 2003 APHHASPDPRAP-----PASADPHRETKSPKPSIQLELRSLCYHSSSYSPGVEV 2056
      |||||
      985 SGHNBERGPPGQGLPQGPGTAGEPGRDNPQ-----SDGQPRDGSFGKGPDR 1033
QY 2057 SPVSSPSTLTHDKGLPKHLELDPKSHLEGELRKPQGPVYKLGGAHMLPHLRPLPSOPSS 2116
      |||||
      1034 GENSGPGA---PGAQH-----FGPRPVPVSGSKSGDRGTGP---AGSG 1073
QY 2117 SP---LLOTPA-----GVKQHORVVTLAGHISEVITQDYTRHPQOL 2155
      |||||
      1074 APGAPAGAPGAPGQPGRDPKGTGERSNGIKGR----- 1108
QY 2156 SAPLEAPLYSPFGACPVLDLRRPPSDLYLPPPDGAP---ARSGHSEGGKSRPERNK 2211
      |||||
      1109 -----GPPG-----NPG---PRGSPGAAGHOGAIGSPGAPGRGVPVGPBG 1145
QY 2212 TSVLGGEGDGL---PVSP-----EGMTEPCHSHSAVYPLLRYRGEQTEPSR 2256
      |||||
      1146 PP---GKDGTSGHGPTLPFPGRNKGKRGSG---SPGH-----FGQGPPEP 1188
QY 2257 MGSKSP-----GNTS-----QPPAFSKLTESNAMSVMKSKOEINKLN 2295
      |||||
      1189 PGAPPCGCGGAAIAGVGEGKSGFSPYGGDDPMDFKINTBEINSLSKSVNGQIESLIS 1248
QY 2296 TH-NNEPEYN 2305
      |||||
      1249 PDGSRKNPARN 1259
Db

```

RESULT 63

```

CGHUTL
collagen alpha 1(III) chain precursor - human
N/Alternate names: procollagen alpha 1(III) chain
C/Species: Homo sapiens (man)
C/Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000
C/Accession: S05272; S04642; E0011; S01726; S04687; A90399; A94562; I51868; S59511; A90
R/Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A/Reference number: S05272
A/Accession: S05272
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1240, 'V', 1242-1466 <PRC>
A/Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuitvanleml, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A/Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma
erences.
A/Reference number: S04642; MUID:89350838; PMID:2764886
A/Accession: S04642
A/Molecule type: mRNA
A/Residues: 1-1196 <ALA>

```

```

A/Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A/Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A/Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (
A/Reference number: E0011; MUID:89378752; PMID:2777083
A/Accession: E0011
A/Molecule type: DNA
A/Residues: 1-176 <BN>
A/Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
R:Tomon, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A/Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A/Reference number: S01726; MUID:88303360; PMID:3405773
A/Accession: S01726
A/Molecule type: mRNA
A/Residues: 1-170 <TM>
A/Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A/Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A/Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
A/Reference number: S04887; MUID:89386015; PMID:2780304
A/Accession: S04887
A/Molecule type: mRNA
A/Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A/Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g30045
A/Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Sever, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A/Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A/Reference number: A90399; MUID:77134724; PMID:557335
A/Accession: A90399
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A/Experimental source: liver
A/Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R:Sever, J.M.
submitted to the Atlas, December 1977
A/Reference number: A94562
A/Accession: A94562
A/Molecule type: protein
A/Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A/Experimental source: liver
A/Note: author submitted corrections to A90399
R:Milewicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Walstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A/Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
fepring.
A/Reference number: I51868; MUID:93304430; PMID:8317500
A/Accession: I51868
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 186-194 <ML>
A/Cross-references: GB:S62925; NID:g386425; PIDN:AD13937.1; PID:g4261637
R:Chodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A/Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A/Reference number: S59511; MUID:96067614; PMID:7487954
A/Accession: S59511
A/Molecule type: mRNA
A/Residues: 302-423 <CH1>
A/Cross-references: GB:S79877; NID:g1195576; PIDN:AA835615.1; PID:g1195577
R:Sever, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A/Reference number: A90414; MUID:79000343; PMID:687591
A/Accession: A90414
A/Molecule type: protein
A/Residues: 399-675, 'N', 677-727 <SEY3>
A/Experimental source: liver
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991

```

QY 2090 QPGVYKGG--BAHLPHLRPLPESQSSPLLOTAPGVKGRVYVTLAQHISEVITQDY 2147
 Db 2540 ---PLNRAGTWKREHSKI-----SSSLPRVSTWRTGSSSSITLSASSESEKAKSED 2588
 QY 2148 TRHHPOQLSAPLPAPLYSFGPASCPLDLRRPPSLDLYPPDPHGAPARGSPHSEGGKSP 2207
 Db 2569 ERH-----VSSMPAP-RQMKENQVPTKGTWKRIKESDLSPTGMASQASSSGAASAEKBP 2642
 QY 2208 EPNKTSVLGGGEGDIEPVSPPEGMTEPGHSRGAVALPYLDGEOETEPFRMSSKSGNTSQ 2267
 Db 2643 L-----LYQMAPPVSKTEEDYVWRIEDCPI-----NNRPSGRSPGTGN-- 2679
 QY 2268 PPAFSTKLTEGSAWVSKKQKQINKKLTNNRNEBNYISQPGTEIFMMPAITGGLMTY 2327
 Db 2660 -PPVIDSVSEKSSSIKOSKD-----SKDTGKQ-----SVSGSPVQVGLSTR 2723
 QY 2328 RSOAVQEHASTNMGLEAIIIRKALMGKYQWESPELSANAFNPLN-----ASASLPAAPI 2383
 Db 2724 LMSFQVAPAPKQKTEA-----KPGQS-----NPVSIATAEICIAERTFP 2764
 QY 2384 TAADGRSDHTLTSPGGGKAKAVSG--RPSHRKAKSPAPGLASGDRPPSVSSVHSEGCN 2440
 Db 2765 SSSSS-SHG--SSPGTVAARVTPFNYPNPSPKSSADSTASAPSQIPVSTNTKKRD-- 2819
 QY 2441 RRTPLTNVWEDRPSSAGSTP 2461
 Db 2820 SKTDITE-----SSGAQSP 2833

RESULT 62

collagen alpha 1(III) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 13-Aug-1999
 C:Accession: S59856; S62120; S16373

R:Roman, P.D.; de Crombrughe, B.
 Gene 147, 161-168, 1994

A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
 A:Reference number: S59856; MUID:95011609; PMID:7926795

A:Molecule type: DNA

A:Accession: S59856

A:Residues: 1-1464 <TOM>

A:Cross-references: EMBL:X52046

R:Roman, D.

submitted to the EMBL Data Library, November 1994

A:Reference number: S62120

A:Accession: S62120

A:Molecule type: DNA

A:Residues: 1-866, 'G', 868-1464 <TOA>

A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322

R:Meisner, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16373

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1442-1464 <MEM>

A:Cross-references: EMBL:X57983; NID:950476; PIDN:CAA41048.1; PID:950477

C:Genetics:

A:Insertions: 29/1; 96/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 29

58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 868/3; 886/3; 940/3; 976/3;

C:Superfamily: collagen alpha 1(III) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-154/Domain: propeptide #status predicted <PRO>

F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>

F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>

F:126-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 2.6%; Score 343.5; DB 2; Length 1464;
 Best Local Similarity 20.5%; Pred. No. 5.6e-05;

Matches 359; Conservative 104; Mismatches 573; Indels 715; Gaps 84;

QY 715 EEMTEAALHASGNEVDRGEC-----SGPAVNNSSDRETSIPSPYTEAKXTGQNGP 767
 Db 64 DDIIIDDEEPLDCPNEELIPFGCCALCPQPSFPAPV-----LPDGH-----GQGP 108
 QY 768 K-----PPATGADAPP-----PGPTPPRRSTRAPV----- 794
 Db 109 KGDGPFCPIPCRNDDPGLPGPGGLPGPPGSGICIESCPTGQNYSPQSDYDVYSGVGM 168
 QY 795 --EPTPASEATGAPTPPAPP-----SPSAPPVVPKKEEETAAAPVEEGEOKP 845
 Db 169 GGYGCPA-----GPPGPPGPPGSSGHPGSPGSPQYQGPBGQAGPAPP-----GP 216
 QY 846 PAEEELAVDTKAEPEVSKSECTEEABEGPAGKDAEAAEAFAEALAEKKEGSGRATT 905
 Db 217 PGA-----LQPA-----GPA-GKQGE----- 231
 QY 906 AKSSGAPQDSDSSATCSADEVDEAGGDKNLLSPRSLLTPGTDP-ANASPOKPLDLK 964
 Db 232 --SGP-----GRGEKGLGPGCIKPGPAGMPGPPGKKGHPDGR 270
 QY 965 QLKORAAAIPIQVTKVHEPPREDAPTKPAPAPPQNLQPSDAPQDQSS----- 1018
 Db 271 NGEKGETGAPGLK-----GENGLPDGNGAPGPMGGRGAPGGRGRLPGAAGRGND 322
 QY 1019 -PRGSRSPAPADKAPAAAOQLPGDPPCWTGSLGPPVPPREVYKASPHAPPSASYS 1077
 Db 323 GARSDGQPGPPG-----PGTAG--FPGSPGAKGEVGPAGSPGNSGP 364
 QY 1078 APPGHPLPLGLHDTARPVLPAPPTISNPPPLISSAKHSVLERQIGAISQMSVQLAHVY 1137
 Db 365 GQGEPPGQAGAGQGP--PGPPG--NNGSPGCKEMKMPAGIPGAPGLIG----- 410
 QY 1138 SEHAKAPVGV-TWGLPLPMDPKLAPPGVGVQOLSPRQAGPPESIGVPTAEASVLR 1196
 Db 411 --ARGPPGAGTAGIPIGTRGSPGEPKNGAKGEP--GARGRGEGASGIGIPGKEDKD 466
 QY 1197 GTALGVGGSITGIGIPSTRVPSDSAITYRSI--THGTPAVLYKGTITRIIGESPERL 1255
 Db 467 GSP--GEPGANGLPQAGGERGSPG--FRQAPGPNIGP-----GEGPPGE 507
 QY 1256 DGRGDSLPKGVHLYEGKKGVLYSEGGMSYTGQSKEDGRSSGPPHETAPKTYDWE 1315
 Db 508 RGGGPPAGPRVABEPGRD-----TPGPGI-----RKMPSGPPGPN-----D 547
 QY 1316 GRVGRATSSASIEGLMGRALPPEHSPPHLLKEQHHRIG-----STTGIPRSVBAQEDYL 1371
 Db 548 GKPPGPGS-----QGESGRPPGPPSPG--RGQGVMGFPPPKGNDCAP----- 589
 QY 1372 RREKTLKREGTPPPPPSRLTEAYKTQALGPLKLRANHGGLVATYKAGRSIHETPRE 1431
 Db 590 --GNNGRGGPPGPPGLPG-----PAGKNGETGP--QGPPTGPPAGDKDGS----- 632
 QY 1432 ELRHTEPLPLAPRLKEGSIQTGP-----LKTDTGASTTSGKQVRSLLGSPG 1481
 Db 633 -----PPGPQGIQGIPTGTGPPGEGKGPKEPPKGVAPAPGPKGD-----SGAPG 680
 QY 1482 RTTPPVHPLDVADARALERACYESLSKSRGTASSSGSITARCAPVTPBELGK-PROSP 1540
 Db 681 ERGP-----PQTGIPICARGAGAP--GPGGKGPAPGP 712
 QY 1541 LTYEDHGAPFPAGHP--RGSVYVTKREPPRLQDESSLSSKASQDRKLTSTREIAKSPHS 1598
 Db 713 GPPGASGPPGLQCPGGERGP--GSPPKKEKEPAGAGAD----- 751
 QY 1599 TVPEHHHPDISPYEHLKGVSGVDLYRSHIPLADPTSIPIGIPLDAAAAYVLPRLHAPN 1658
 Db 752 -----GVPGKD-----GPGG----- 761
 QY 1659 PTPVHLYPPVYLIRGVPDTALENROTIINDYITSQMHNTATYMAQPADMLRGLSPRES 1718
 Db 762 PAGP-IGPP-----GAGCGDGRGEGGSP----- 784


```

Db 161 ESCEKLECEGGERPQOOPHRAORSPOOQPSRLHRPQNEPQODPTPEEETILE 220
Qy 720 EA-----EALHAGNEVPRGECGPATVNNSSDTESI:PSPHTEAKOTGONGPK----P 769
Db 221 SSLPLPLEEAH--GPRGLKGEKGPVAVLEPOMLEVGPGR-----EGPAGLIGP 268
Qy 770 PATIGADGP--PPGPPTPRRTSRAP:IEPTPASEATGA:PTPRAPPSAP-----818
Db 269 PGIQGNPBPVQDPGERGPPRGAG-----LPQSD--GAPGPCT--SLMLPFRFGSGGD 318
Qy 819 -PPVPRKEKEKEE-----TAAAPVVEEGBEQKPPAAEBLAVDTGKAEP---VKSE 865
Db 319 KGPVVAAGEAQAQALIQARLALRGP:PGPMGYTGR-----GLQDQSGSGLKGE 368
Qy 866 CTEAEAEER-----AKGDAEAAETAGALKAEKKGSGGATTAKSAGAPQSDS 917
Db 369 SGDLGPQGPGRPOGLTGLGAGRGRAGPQDARLTLD:PGVKGRGFDGL:RGLRGEKXH 428
Qy 918 SATCSADSV-----DEAGGDKNRLSPRPSLL:PTGDPANASPOKPLDLKQLKQRAAA 972
Db 429 RGDTPRGRLPGRPGEDGERGDDGEI--GPRGLPGESEGRGLLGPGR-----473
Qy 973 IPRITQVKNHBPREDAAPTKAPAPRPPQML:QESDAPQOPGSS--PRKSRSPAPPA 1030
Db 474 -----PGIRGPPGVRCM-----DGPQPKSLGPQGB--PGPRG 505
Qy 1031 DKEAFABAOKLPGRPPCMTSGL:PPVPPREVITKASPHAPDASASTAPRPHPLPLGLHD 1090
Db 506 Q-----QGTFG-----TOGLPGR-----QGAIRPHGKKGQ--532
Qy 1091 TAPVLP:RPRTISNPPLISSAKHPSVLE:ROIGALSQGSVQLAHVSEHAKAPVPTM 1150
Db 533 --KGLPMPGSDGP-----GHPG-----XKGP--PGTK 558
Qy 1151 GLPLPMDPKCLAFPSGVQEOQL:SPRGQAPRESLGVPTA:QEAASVLRGTA:SGVSPGSITX 1210
Db 559 GKP-----GSPGQGPGLGVPGRQGVKGVNG--IRGLKHKHKEK 594
Qy 1211 GIPSTRVSDAITYTGRSI--THGTADVLYKGTITRII:GEDSPSLDGRBESL:PKGHVI 1269
Db 595 G-----EDGPPGKGDIGVKGDRGEVGPSS--RGEDEPGE-----PKGRITG 634
Qy 1270 YEGKKG--HVLSEGGMSVTQCSKEDGRSSGPRPHETAAPKTYMMEGRVGRATISASI 1327
Db 635 PTDGPPRGGLMGEKGLGVPLG:PYRGR--QCPKKSGLRP-----GPRC-ASGEKGA 683
Qy 1328 EGLMGRAP:RERHSPHHLKEQHHRIGSI-----TOGIPRSYVEAOE 1368
Db 684 RGLSGKSGPRGRGPHSGTQGRGPRGATGKSGAKGTSGDDPHGPRGREGLLGP-----737
Qy 1369 DYLRREAKLLKREGT--PPPPSRDIT:EA:YKTQ-----ALGPIKLKPAHEGLVA 1416
Db 738 -----SGSNGPFGKGPGRPGKGLP:PHQQRSEVGRQKKGK--PGPRGVVG 784
Qy 1417 TVKEAGRSIH:IPREHLHTPRL:PLAPRLKESITOGPTLYUDTGA:STGSKKH:DVSL 1476
Db 785 PQAGAGET--GPMERGH--PGPRGRB--QGLP-----GTAG--KCGTKGD 824
Qy 1477 TGSPP-----RTPPVHPLDVMA:DA:RALERACV:ESLKS:RPTAS:SG--GS:IA:R 1524
Db 825 PGPFGAPGKDPAG:RGRFGRGL:PGTAGGRLK--NEGSGPRGAPGSGERGAAGS 881
Qy 1525 GAPVIVBELGKRGROPLTYEDHGAPFAGHLPRGSPVYTMREPRPL:QESLSSKASQDRK 1584
Db 882 GGP--IGRGPRG-----PGPGGAAGEKGVPEKGPITG-----913
Qy 1585 LTSTREIAKSPHSTVPEHNP:IPYEHLLRGVGVLDYRSH:ILAEDPTSI:RPIGLPD 1644
Db 914 --PTGRDVGQGVGL-----PGRAP--PGVAGEBDGKEV-----GVGEQ 950
Qy 1645 AAAAYVLPRLA:PNPTVPLYPYLIRGYRPTA:LENRTI:INDYITSGQMHNTATAMA 1704
Db 951 KGTGNKGEHGPGRGP-IGPV-----GQPGAAGDGAAGL-----GTQGHFG-----A 995

```

```

Qy 1705 QRADMLRGLSPRESSLALNTVAGPRGII:LDQVPHLPLV:PTPCTPATAMDBLAVLPTA 1764
Db 996 KADGETGFNGP:PPPTIGL:GIPSG--EKETEDVGMPGPP-----PGP 1040
Qy 1765 PQPSSHHSSSPLP-----GPTILTKPTTSSSRERDRDREREREKSI:LYTS 1817
Db 1041 RGPAGPNCADGPQGP:PGVGNLGR:PGKEGEGSGSGCIOGEPVKGRGRGKGG--ES 1096
Qy 1818 TTYVHAPVWRPTEQSSGSSGSGGG--GSSSRPASHAHQHSPI:PTDOLQORP 1875
Db 1099 GQPEPGR--PGAKGQGDGDPKGNP:GVFPEDPGPRGK-----GPRGODGAK--1146
Qy 1876 SVLHNTGKGIITV:VEPSKPTVLR:STSSSVYRAATPRP:THPGLGGTLGVPTLMEP 1935
Db 1147 --GDRGEDG-----EPQRP-----GSGPPTGENGP--PLG-----1174
Qy 1936 VILPKEAPRVARPERPRADTGAFLAKPRASGLEPASPEKSGEPRLVPVSGHATIA 1995
Db 1175 -----KRG--PAGSP--GSEGR-----QCGKGA 1194
Qy 1996 RTPAKNLA:PHNASDP:PPAPASADPHREKTQSKPFI--QELRLSLCYHSSYSPEVE 2054
Db 1195 GDPGALGAPGKTGEVGPAGPAGKPRP--DGLRGLP:GVSQCGRRPGATQAG--PPG--1246
Qy 2055 PVSPPVSSPLTHDKGLPRHLELDKSH--LEGELRPKQGVNKLGGEAHL:PHLRPLPES 2112
Db 1247 PVGPGLPGLRGDAG--AKGEKHGDLGLGLP--PGQGEKGD-----RGLGR 1292
Qy 2113 QPSSPPLQTPARGVGHORVVTTLAQHISEVITODYTRHHPOLLSAPLAPLYSPGASCP 2172
Db 1293 QGS-----PGQKXGEMGI-----PGASGP 1310
Qy 2173 VLDLRPPSDLYL:RPPDHGAPARSGPSHSGKSRPEPNKTSVLOGGEDGIFVSPBEGMT 2232
Db 1311 I-----GPGG--PP--GLGPAGPKAGKATGP-----GPRGKGVQGP--1348
Qy 2233 EPGHSRAVYPLVLRDGEOTEPRMGSK-----SPGNTSQPPAPFESKL 2275
Db 1349 HPGPGEVITQPLPTQMPKTRRSYVDGSLKMOEDBALPTGAGSPGGLLEIFGSL 1403

RESULT 61
149505
adenomatous polyposis coli protein - mouse
N:Alternate names: APC
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: 149505
R:Su. L.
Science 256, 668-670, 1992
A>Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
A:Reference number: 149505; MUID:92263101; PMID:1350108
A:Accession: 149505
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2845 <RES>
A:Cross-references: GB:M68127; NID:g191991; PIDN:AMB59632.1; PID:g191992
C:Superfamily: adenomatous polyposis coli protein

Query Match 2.6%; Score 344; DB 2; Length 2845;
Best local similarity 19.0%; Pred. No. 0.00011;
Matches 516; Conservative 334; Mismatches 1073; Indels 798; Gaps 124;

```

```

Db 794 -----PGRGVK-----ADGIRLKTGK-----814
QY 1614 LTRGVGVDLVRSHPLEADPTSI-----PRGI-----PLDAAAYVLPRLNPNPTVPHL 1664
Db 815 --KGEDEFGFKDMKIDKDRGEIGPPRGEDGPPRGPPGDDPEPLPTBEKGL 872
QY 1665 YPPYLIRYPTDALENROTINDYITSOQMHNHTATAMAQADMLRGLSPRESSLAANY 1724
Db 873 GVFGVLP-GRY-----GRQ-----884
QY 1725 AAGRGITLDSQVPHLPVLPPTPGPATAMRLAVLPAPQPFSSRHSPLSGPPTH 1784
Db 885 --GPKSGIFP-----PGFPMANGKGR--CTPKPFGPRGKGTGREGRGPPG 930
QY 1785 LTKPTTSSSERERDRDRDRDREREKSLITSTTVEHAPLWRPTEQSSGSSSG--1842
Db 931 IT-----GKPGPKNGSGDGPAGPP 950
QY 1843 GGGGSSSRASHAHQHSPISTPTDALQORPSVLANTMGKITAVPSKPYLKRST 1902
Db 951 GERGPFGPGPTGPPGKPPGPPGDLPGHGGRGFTGFG--K 994
QY 1903 TSSPVPAATFPATPCPLGTLGVYPTLMEPVLLPKCAPVAPRPERPADTGAFLAK 1962
Db 995 TGPPGPPGVVGPQG--PTGET-----GPMGERGHPGPG 1026
QY 1963 PPARSGLE-----PASPSKSGSE-----RPLVPVSGHATTARTPAKN 2001
Db 1027 PPEQGLPGVAKEGTKGDPGAGLPGKOAPGLRGFPDRLGPGV--ALG 1077
QY 2002 LAPHASDPBPAP-----PASAD-----PIRETKSKPSIQLELRSLGHSSTS 2049
Db 1078 LKSGSEGPBPBPAGSGPGRGPAAGPIGIPGRPPGPBPAG--KVPGEK-G 1131
QY 2050 PEGVE-----PVSPSPSLTFHDKLPKHLLELDKSHLEGELRP-----2089
Db 1132 PQGPARDGLQGVGLPGRAPGVPPGPDGDKELBEPQKSGKDGKGGPPGTPGQ 1191
QY 2090 ---QCPVKLGEAAHLPHLRPLPESQ-----PSSPLLOTAPGVKCH 2129
Db 1192 PICQPPSGADGE-----PGRGQQLFGQKDESGKSGPPGPPVGLQGLPPGPB 1243
QY 2130 QRVVTLAQHSVITQDYTRHHPQLSAPRLVLPFGPASCVDLRRPSPVLP--2186
Db 1244 KGB-----TGVGQWGP--GPPGPPGSAAPADGP---QBPGGIGNPGAV 1286
QY 2187 ---PPDGAPA-----RGSPPHSEGGKRSPEPN 2210
Db 1287 GEKGEFGEAGEPGLPBGEGPLGKGRGKGEAGPAGAPPPKPPGDDGPKSGPP-1345
QY 2211 KTSVLGGEGDGLPVPSP-PRGMPTEPGHSASVPLYLRDGEOTEPBRMSKSGNTSQP 2269
Db 1346 ---VGFPPGPPGPPGPPGPAQDGPDPGDK-----GDDEPQGTG--SPGPPGE-1388
QY 2270 AFESKLTSSSNAWKSKQKQINKKLTNHNNEBENYISQGTJIFMMPLITGMLTVS 2329
Db 1389 ---GSGPPGKRGPPG-----PA-----GPGS 1407
QY 2330 QAVOEHA STMGLFAIIRKALMGKYDOWESSPLSANAFNPLNLSA-----SLPAM 2381
Db 1408 ROEKAKAGKAGL-----EGPPGKTPIGIPGAPGKPPGDPGLRGIPGV 1451
QY 2382 PTPAADGRSDHTLTPSGGGKAVSGRP-----SSRAKSPAPAGLSGDRPPVSVS 2433
Db 1452 GEOGLPQ-----SPGPDPPPGPMGPPGLPGKJDSGPPGKGGHGLGLTGPPEQG-1503
QY 2434 HSGGDCMRRTPLTNRWEDPSSAGSPRPYPNPLINLQGWASPPPPGLPAGSGPLAG 2493
Db 1504 -EKGD-----RGLPFGSSG--PKGEGGIGTGSGLPFPFPGLPFPBPBKKA 1549
QY 2494 PHHAMDEEPK 2503

```

```

Db 1550 KGSSGPTGPK 1559
RESULT 60
COH2E
collagen alpha 2(XI) chain precursor - human (fragment)
N:Alternate names: procollagen alpha 2(XI) chain
M:Contains: proline/arginine-rich protein (PARP)
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C:Accession: S34790; A32645
R:Zhickova, N.I.; Brewton, R.G.; Mayne, R.
FBS Lett. 326, 25-28, 1993
A:Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage
      brain.
A:Reference number: S34790; MUID:93314796; PMID:8325374
A:Accession: S34790
A:Molecule type: mRNA
A:Residues: 1-663 <ZHI>
A:Cross-references: EMBL:L18987; NID:G306439; PIDN:AAA5498.1; PID:G306440
R:Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;
J. Biol. Chem. 264, 13910-13916, 1989
A:Title: The human alpha2(XI) collagen (COL1A2) chain. Molecular cloning of cDNA and gc
A:Reference number: A32645; MUID:89340485; PMID:2760050
A:Accession: A32645
A:Molecule type: DNA; mRNA
A:Residues: 586-1546 <KIM>
A:Cross-references: GB:J04974; NID:G180714; PIDN:AAA52034.1; PID:G180715
A:Note: parts of this sequence were determined by protein sequencing
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL1A2
A:Cross-references: GDB:119788; OMIM:120290
A:Map position: 6p21.3-6p21.3
A:Introns: 1302/3; 1320/3; 1332/3; 1350/3; 1440/1; 1477/3
A:Note: the list of introns is incomplete
C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CGH
3(XI) chain (see PIR:CGH06C), initially linked by disulfide bonds among their carboxyl-
      imed with desmosine cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cel
A:Note: may play a role in controlling the lateral growth of collagen II fibrils
C:Superfamily: collagen alpha 1(IV) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F:1-254/Domain: non-collagenous (fragment) #status predicted <NC3>
F:1-187/Product: proline/arginine-rich PARP protein (fragment) #status predicted <PARP>
F:255-305/Domain: collagenous, triple helix #status predicted <COL2>
F:306-342/Domain: non-collagenous #status predicted <NC2>
F:343-1356/Region: helical
F:429-431/Region: cell attachment (R-G-D) motif
F:447-449/Region: cell attachment (R-G-D) motif
F:1257-1259/Region: cell attachment (R-G-D) motif
F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide
F:1381-1546/Domain: carboxyl-terminal propeptide (fragment) #status predicted <CTP>
F:1403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status aty
F:109-163;1511-1545/Disulfide bonds: #status predicted
F:109-163;1511-1545/Disulfide bonds: #status predicted
F:126/Modified site: 5-hydroxylysine (Lys) #status predicted
F:126/Modified site: 5-hydroxylysine (Lys) #status predicted
F:126/1266/Binding site: carboxylate (Lys) (covalent) #status predicted
F:327;933;1008;1017;1035;1038;1296;1305;1317;1320/Modified site: 4-hydroxyproline
F:329/Modified site: 4-hydroxyproline (Pro) #status atypical
F:342;1023;1299/Modified site: 5-hydroxylysine (Lys) #status experimental
F:342;1023;1299/Binding site: carboxylate (Lys) (covalent) #status experimental
F:1427;1433;1450;1459/Disulfide bonds: interchain #status predicted
F:1427;1433;1450;1459/Disulfide bonds: interchain #status predicted
Query Match 2.6%; Score 344.5; DB 1; Length 1546;
Best local similarity 21.2%; Pred. No. 5.4e-05;
Matches 367; Conservative 122; Mismatches 631; Indels 615; Gaps 92;
QY 664 ETLQGHKLMF---KERNARKKKKQAPAAASEAAFPVVEBEBEAGS-VGNGNEEWE 719

```


Db	1498	PDVADPTSSKQSKRSIRKVDVEEFPALNKRTPSAGKAMHTPKPAVSGEKNIYAMQCT	1557
Qy	1168	KOEOLIS--PRGAGPPEISLGVPTAOEASVLTGALSGVPGSGITKGIPISTRVPSDAITYR	1226
Db	1558	PVQKLDLIENTLGSRRRLQTPKE-----KQALIEDLAG---FELPQTRGHTEB-----	1603
Qy	1227	GSITHGTPADVLYKGTITRIITIGEDSPSLDRGDEDSLPGHVIYEGKKHVLSTEGMSV	1286
Db	1604	-SMNDDTKAKVACKSSQPD--DKNPASSKRRLTSLQK-----VGAYEBLLAVG---KL	1652
Qy	1287	TQSGKEDRSSGGPHNE-----TAAPKRTYDMMWEGVGRALISASIEGLMGAIPEPR	1339
Db	1653	TQSGEITHTEPEPGDGSKMAFMESKQILDSASITG---SKROIPTPKSGSEYVED	1709
Qy	1340	-----HSPHHLEQOHINGISIT--OGIPRSYVEAOEDYLAREAKMLKKEGTPPPPP	1389
Db	1710	LAGFIELEFQTSHTKE-----SMINEKTTKVSTRASQPD-----LVDTPISSKQF	1755
Qy	1390	SRDLTEAVKTOALGFLKLPRAHEGIVATVKE--AGRSIHEIPREBELNHPDELPLAPPL	1446
Db	1756	KRSIRKA-----DTEEEFLAPRKQTPSAGKAMH-----TPKPA	1788
Qy	1447	-----KEGSIITQCTPL-KYDTGASITGSKH-----DVSRLGSGGRTPRP	1486
Db	1789	VGEEDKINTFLGTPVQKLDQPGNIPGSMNRLOTRKEKAQALEBELTGFEFLFQTCPTDNPT	1848
Qy	1487	V-----HPLDVWADA-----PALERACYEESILKSRPGTASSGSGSIRAGA	1526
Db	1849	ADEKTTKILCKSPSDPADPTPTTKQRPKPSLKKADVEEFLAPFKLTPSAGAMHTPK	1908
Qy	1527	PVIIVE-----LQKRPSPLYTEDHGAPFAGHLPRGSPVTMBEPTPLQEGSLSSKAS	1580
Db	1909	AAVGEKQINTFVGPHVEK-----LDLIGNLP-GS--KRRQTPK-----EKAKAL	1951
Qy	1561	QD-----RKLTSTP-----RELA-KSPSHVPEHHPHPSPEHHLLRGVSGDIL	1623
Db	1952	EDLAGFKFLFQTPGHTESMTDJKITEVCSKSPQ---PDVPKTPSSKQRIKISLGKGV	2008
Qy	1624	YRSHPIPLA-FDPTSIPIRGIPLDAAAYVLPRLHAPNPYLPPLPYLIRGYPTAALENR	1682
Db	2009	KEEVLVPVQKLIQTS-----GKTYQTH-----R	2030
Qy	1683	QTIINDYITTSQOMHNTATAMAQRADMLRGL-----SPRESSILANVAGRGIIIDLSQV	1737
Db	2031	ETAGDGG--KSJIKAFKESAKOMLDNPANYGTGEMBRWPRTEKBAQSLIEDLAGFK--ELFQI	2085
Qy	1738	P-HLPPVLVPRTPGATATMDRLAYLPTAROPRSSHSSPLSPGCPHTLTPCTTSSSR	1786
Db	2086	PDHT-----EESTDDTKTKACKSP-----PESMDTPST--R	2118
Qy	1797	ERDRREDRDREREREKSI--LTSYTYVEHAP-----IMRPQTEGSSGSSSSGGGG	1845
Db	2119	RRPKTPPLGRDIVELSLAKOLITQTHTDKYPGEDBKINVFRKPAQKLDPAASVYG--	2176
Qy	1846	GSSSRPASHAHQHSPISPRTODALQORPSVLIANTGMKGII-TAVEBSKPTVLRSJSTS	1904
Db	2177	-----SKQRPRTPKKAQRLIEDL-AGLKEFLQFQRPVCTDKDPTTHEKT-TK	2218
Qy	1905	SPVPRPAATFPRAITHCPLOGLDGVPTLMER-----VILPPEARVVARP-	1948
Db	2219	IACSPROPDPVGT-----PTIFKQSKRSIRKADVEEBSIALRRRTSVGVGM	2266
Qy	1949	ERPRADTG-----NAFLAKPARSGLEBPASSKSGSEPRPLVP-----PVGHNATIA	1995
Db	2267	DLPKAGGEDCKMAFKMGTPVQKLDL--PGNLP--GSKKWPQTPYKKAQALIEDLAGFKELF	2323
Qy	1996	RTPAKN-----LAPNHASPDPPAPRPAASDPNREKTOQSKPFSIQE--LELRISLY	2043
Db	2324	QTPGIDKPTTDEKTKIAKCSQORPVDTPASTK--QRPKNIRKADVEEBFLALR-----	2377
Qy	2044	HGSSYSPEGVBEVSVSSPSLTHDKGLFPHLEBIDKSHLEGBLRKQRPYKULGGEAAHL	2103
Db	2378	---KTPSAGKAMD--TPRPVAVSDEKININTFVET-----PVQKLDLGNLT	2417

```

QY 2104 PHLAPBPSSQSSPPLIQTAPGVK-----GHQVNTVLAHISVITD-----YTRH 2150
Db 2418 PGRKQOPOTPEKKEALEDLVGFKEFLPOTPEHTEESMTDKITVESSCKSPQESFKTSRS 2477
QY 2151 HPOOLSAPLPAVLVSFPQASCPVLDLRPP-----SDLVLPDP-----HGAPARQSPHSEGG 2203
Db 2478 SKORIKTIPLVK-----VDMKEEPLAVSKLIRTSSETTQITTEPFGDSKSLKAF 2525
QY 2204 KRSPB---FNKTSVLGGGED-----GIEP-VSPBEGMTEGHSRAVYPLLRYDGE 2250
Db 2526 KESPRIQILDPAASVATGSRRLQRTREKERALEDLVDFKELPISARGHTEESW-----T 2577
QY 2251 QTEPSRMGSKSPGINTSOPPAFFSKLTESNSAMVSKQOEINKKUNTNNRNEPEINISQPG 2310
Db 2578 IDKMTKIPCKSP-----PPELTDVATSTKRCPTKTRPKVEBELSAVERLT--QTSQOS 2629
QY 2311 TEIFMPLRITGLMTRYSAQ-----VOEHAS----- 2337
Db 2630 THTHKEPPASGEGIKVLKQRAKKKPNPVEEPPSSRRPPAPKEKQPLEDLAGFTELSETS 2689
QY 2338 --TNMGLEAIRKALKMGKYDQWE--ESPPLSANAFNNPLNASL-----P 2378
Db 2630 GHTQESLTA-----GKATKIPIGESPLAV-----VDTTASTGRHLRTRVQKVQVKEEP 2737
QY 2379 AAMPITAADGRSDHTLTSFGGGGAKAVSGRPSRKASPAFGLASGDRPP 2428
Db 2738 SAVKTFQSGETTDADKEPAGEDGIKVLKKSAAQOTAPPAASVSGSRPP 2787

```

[illegible]

Db	1100	GSPNIG--PSSPGLPGEKGDGLPGLDGVGVG-----	1133
Qy	2154	QLSAPLPAFLYLSFPGASCPLVDLRRPPEDLYLP-----PFDHGAPARGSPHSEGGKSPK	2208
Db	1134	--EAGLPET-----PQPTSPAGQKGPSPD--GI PSAGEKGEQVLPGRKGFPPFPBSK--GDK	1185
Qy	2209	PNKTSV-----LGG-----GSDGIDIEVSPSEGTTEBGRSAVYPLLYRDE-----QTE	2253
Db	1186	GSKEVEVGPGLAGSPHGI PGVKGEGOF--MGPRPGQGGYGLGTGPHVPGKGBRGRPGQG	1243
Qy	2254	PSRNGSKSPGTSQPPAFPSKLTESNSAMVSKKQELN--KLTATHNRNRPYNI SQPTE	2312
Db	1244	PGLPGHPGP--MGPPGF-----PGLNGPKGDKNQGWPF--GAGGVP	1280
Qy	2313	-----IFNNPATITGTMTRYRQAQVEHASTNMGLEAII-----RKALWG-----KYPDQW	2357
Db	1281	GPKPDPGFGQGPFGIGGSPGIT-----GSKGDMLPGVGGFGQGLPGLQGVKGGQG	1332
Qy	2358	EESPEPLSANAFNPLMASASLPPAMPITAADRSHTLTPSGCGKAKVSGRPSRKAKSP	2417
Db	1333	DQG-----VPGPKLQGPFGPGPYDVYIKG	1357
Qy	2418	APGLASGDRPPSVSSVHSEGDCCNRTPLTNY-----MEDRPSSAGST--PF--PYN	2465
Db	1358	EPGLPGPGPGPLKCOLQGPGRPKGQGVTSVGLPGPGVGPFGDAPQCKGETPFGPGP	1417
Qy	2466	PLIIRLQAGVNASPPPGCLPAGSGPLAGP-----HHAMDEEPKFLIC	2507
Db	1418	P-----RGPGPPGPDGLPGSMGPPGTPSVDHGLVYTRHSQTDDP---LC	1460

```

Db      558 SGEKSSGEIHEVEVKAQSLVTSPPAPSPKPTFVASDQRRKSTAPASSKSCOTEVPKR 616
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      245 LEGLQPOVELPYNOPSDPTQYHENIKINQAMRKILLIYFKSRNHA-----RKOMKOK 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      617 ---GGERVATCLOKQVKSISRQHDLO-----MICKSRSGASSEANLIVAAKSMADV 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      298 FCOBYDQLEALKEKKVERIENPPRRAKESKVREYKOPPEIRKQELQERMOQSHVQR 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      665 V-----XLGAKQOTQTVIKHQPOK-----SMNKQRQRPATPKKPVGEV 702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      338 GSGLSMAARSEHEVSEIITDGLSEBENLEKQKROLAVIPMLYXDQOQIKITNNGLMA 417
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      703 HSQFSTGHANSP---CTIIIGKATHEKHVPARPYRVLANFI---SNQCMDFKEDLSGIA 756
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      418 DPMKYKROWA-----NMMSQEK--ETPREKQMOPKNGGLASFLER-- 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      757 EMFKTPVKEQPOLVSTCHIAISNSENLQKQOQCHDSGEBPLPPTSESFGVFPBAQNA 816
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      461 -----KTVAECVLYYLYLTKKNENKSLVRSYRRRGKSQOQOQOQOQOQOQOQPMWR 514
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      817 AKQPSDKCSAPPLRQCIRREKGNVAKTIPRNTYKNTLSL-----TK 857
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      515 SGOEKDEKEKEKEAKEKEBEKEVENDYEDLLIKETDTSGEDNDEKAVASKKRTANS 574
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      858 TSDTETESKTVSYVNRSGRSTEFNNIKQLPVESKSSEFTNTE---IVECILRKGGQATULL 914
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      575 QGRKRGRTTRSMANANSEBALTPOOSALMASMELNSESRTMEEBEMAKKGLLEHGRWN 634
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      915 QQRREGE-----KEIERPEETYK--ENIELEKEN---DEKKAKMR-----SRTW 954
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 57

[illegible]

cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schuster, C.; Duchow, M.; Wohlsenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67; a very large, ubiquitously
ins
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <Kin>

```

QY      691 SE-EAAPPVEDEMEAS-GVSGNEEMVEEAALHNSG-----NEVPRGECSPATVN 743
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      1010 LQPEINTPTHTKQOLKASLGVGKEBELAVGKFTRTISGETHTTHREBAGDOKSIRTFK 106
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY      744 NS-----SDTESI-----PSPHTEAKD--TGONGP 767
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      1070 ESPKQILDPAARVNTGMKKMKPRTPKREAGSLBLAGFKELFQTPGPSEESMTBCKTIKAC 112
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY      768 KEPATLADGDPPEPPTPRTSRAPIE-----TPASEATGATPPPAAPPSAP 819
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      1130 KSPPESVDTPTSTQWPKRSRKADVEEELALRLKLTTP-SAGKAWLTREPA----- 118
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY      820 PVVPEEKEEBEETAAPVE-----EGEBCPPAAEBELA-----VDTGKA 858
      : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      1181 --GGDEKIDKAFMGTPVQKDLBAGTLPGSKKQLOTPPKKAQMLBLAGFKELFQTPGHT 123

```

Query Match	2.6%	Score 348	DB 2	Length 3256
Best Local Similarity	18.7%	Pred. No. 9.2e-05		
Matches 537; Conservative	352;	Mismatches 1129;	Indels 852;	Gaps 133

Db 1238 ELVAGKTTKIPCDSPQSDPVDTPPTSTKQRPKRSIRKADVEGELLACRNLMPSAGAMH 129

```

OY      4  STOLVAQWRAATERERYPHSLSYQVQIARTHTDVLLEYQHSNR-VAS----- 51
Db      325 SVQTPSPKAVGASFPIYEPARKTKTPVOYSQOONS-----DQKHKNKDLVYTTGREGSEVNIKGS 380

OY      52  -----HLSPGSIIOFORRRPSLL-----SEFOQGNERSQOELHLPRESHYLP--EL 95
Db      381 EGFKAGDKTLIPRK-LSTRNRTPAKVEDADAASATKREMLSKTKGSIITDVEVLPTETEI 439

OY      96  GKSEMEPIESKRPRLELLPDLPLRPSL-----LATQOPAGS-----EDLTOKRSI 141
Db      440 HNEPFLITMLTQVENKIKQDSLSKPEKLGTTAGQWCOSLPGIASSVDINNFGDSINESGCI 499

OY      142 TGKLEPVSPPSPPHNDPEL--ELVPR-----LSKEELIOWMDRVDRBITWVEQOI 190
Db      500 PLKRRRVV--RGHRLPELPDENLPNNPLPKRGGAAPTKRKSIVMHTPPLVKKIIKEQOP 557

OY      191 SKLKKKQOOLEEB-----AAKPRPEKVPSPPIESKHSRLVOIITYDENRKAEAARI 244

```

```

Db      1298  TPKPSVGEKDIITVCGTPVQGLDTLNTLTSKRR-----PQTRK 133
Qy      962  DLKOLKORAAALPIQVTKVNEHPREDAAPIKAPAPAPPONLOPESDAPOQPGSSPRG 1020
Db      1338  BEAQLAEDLTGPKELFQTPGHTBEAAVAGKTKMPCSSPPES-----ADTPTS 138
Qy      1022  KRSAPAPADKAPFAAE-----AQLPBDPPCMTSGLPFPVPREVIXA 106
Db      1387  TRQOKTPELKEKDVOKELSALKKLTQTSGETTHTDKVPGEDEKSIAPF-----KRTAK 143
Qy      1066  SEPAADPSAFSYPAPGHP---LPIGLHOTA-----NPVLPRPTISNPRLISSAKH 111
Db      1440  --QKLDPAASVTSGRKHPTKEKQAQRLIEDLAOMKELFQTPVCTDKTTHBKTKIKICRSQ 149
Qy      1115  PSVLEIQTGAISQSGMSVOLHVFSE-----NAKAPVGVPTMGLPLP--MDPKLAFPSGV 116

```


Qy 1479 SPGRTPPVHLDVMDADRALEACYSLSKSRP-GTASSSGSGTARGAPVIVPELGR 1537
 Db 1353 GMGQWLPAASGADPLCRNPA-SRSLKGLNLSKSLAADCPKAEAFNSPELPSLGRAR 1411
 Qy 1538 QSPFLYEDHG-----APPAGHLPR-----GSPVTRMEPPRLQ--EGSLSSKASQDRKLT 1586
 Db 1412 -APKRSKSKGTGTLGPKGFLERKPCGQPLL-APHDRASSVQGGEDSSGGGKK-- 1467
 Qy 1587 STPREIAKSPHSTVPEHHPHDIPSYEHLLRGVSGVDLYRSHIPLAFDPTSIKGIPLDAA 1646
 Db 1468 -PKTEELGPASQPEGRP-----CQOTRAQKPGQAS 1499
 Qy 1647 AAYLPLRLAENPYTPHLYPYLLRGYPTALENRQTIINDYITSQOMHNTATAMA-- 1704
 Db 1500 YSSYSKR-----KRLSRGRKXTA-----HASPCKGRYR 1528
 Qy 1705 QRAMLRLGLSPRESSIALNVAAGPRGIDLSQVPHLPVLPPTPTPATMDRLAYLPTA 1764
 Db 1529 RROQOVLPDLPAPEIRLKIYSSCKRLADSRTP-----A 1563
 Qy 1765 POPFSSRRHSSPLSPGPTHLTKPTTSSSRERDRDRDRERREKSIITSTTV-- 1821
 Db 1564 FSPFV-----RVEKDAYTITICTVNS 1585
 Qy 1822 ---EHAPWRPCTEQSSSSGSSG---CGGSSSRPASHAHQHSISPRTQALQORP 1875
 Db 1586 PGDEPKPMPKSSSAASSSTSSSLPAGASLITFPFG-----SVLQGRP 1630
 Qy 1876 SV-LHNTGMKGILITAVEPSKPTVLASTSTSPVPAATFPATPCLGTLGVDVPTLME 1934
 Db 1631 SLPLASTMHLG-----PVYSALSTSLCY-CCLCONPAMPDLGLDLCGPYTP---- 1676
 Qy 1935 PVLLPEKAPRYARPERPRADTGHAFIAKPPARSGLEPASPSKSGSEPPRLVPSGHATI 1994
 Db 1677 EHCLPKKKPKLKEKARLEGTLEASLPLERTLKGLKESASATTA--PTATITTPPAL 1733
 Qy 1995 ARTPAKNAIPHASDPPAPAPASADPHREKTKQSPFISQELHLSLGHGSSYSPEVE 2054
 Db 1734 GRL-----SRDGPADPAK-OGPLRTSARGLSRRLQ-----SCYCCDG-- 1770
 Qy 2055 PVSPVSPSLTHDKLPKHLLELDKS--HLBGLRPPKPGPVKLGGAHLPH 2105
 Db 1771 -----QGDGQ--EEVQADSRKRGCKSKAPTEPG-----GDTEHWH 1807

RESULT 56
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
 C:Accession: A33525; S01454; A28066; A02864; A25636; A2301; S19079; A32003; A31766; S19
 R:Mutlukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff
 A:Reference number: A33525; MUID:89197932; PMID:2703490
 A:Accession: A33525
 A:Molecule type: mRNA
 A:Residues: 1-1669 <MUT>
 A:Cross-references: EMBL:J04694; NID:9556296; PIDN:AAA50292.1; PID:9556297
 R:Wood, L.; Thieriault, N.; Vogel, G.
 FEBS Lett. 227, 5-8, 1988
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al
 A:Reference number: S01454; MUID:8811221; PMID:3338568
 A:Accession: S01454
 A:Molecule type: mRNA
 A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'P', 404-480, 'L', 482-492, 'H', 494-71
 A:Cross-references: EMBL:X06777
 R:Killem, P.D.; Burbejo, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha
 A:Reference number: A28066; MUID:88243724; PMID:3379041
 A:Accession: A28066

A:Molecule type: mRNA
 A:Residues: 1-129 <X11>
 A:Cross-references: EMBL:J03758; NID:9192669; PIDN:AAA37439.1; PID:9192670
 R:Oberbauer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogel, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1669 <OBE>
 A:Cross-references: EMBL:X02201; NID:950233; PIDN:CAA26132.1; PID:91333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zou, G.; Vogel, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
 A:Reference number: A25636; MUID:86301886; PMID:33755692
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:9192286; PIDN:AAA37340.1; PID:9192287
 A>Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Finlay
 J. Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
 A:Reference number: A94680; MUID:87250460; PMID:3597383
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:9192282; PIDN:AAA37340.1; PID:9387115
 R:Killem, P.D.; Burbejo, P.D.; Martin, G.R.; Yamada, Y.
 J. Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
 A:Reference number: S19079; MUID:88315019; PMID:2842328
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:9192673; PIDN:AAA37442.1; PID:9466503
 R:Kayes, P.; Wood, L.; Thieriault, N.; Kurkinen, M.; Vogel, G.
 J. Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:8906738; PMID:3196626
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAY>
 A:Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37437.1; PID:9450449
 R:Burbejo, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promo
 A:Reference number: A94220; MUID:89071759; PMID:3200851
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M23333; NID:9340878; PIDN:AAA51625.1; PID:9535668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J. Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
 A:Reference number: S19094; MUID:86196099; PMID:3009468
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
 A:Reference number: S16909; MUID:80246483; PMID:6772473
 A:Accession: S16909
 A:Molecule type: Protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
 A:Reference number: A25991; MUID:82186723; PMID:6804236
 A:Accession: A25991
 A:Molecule type: protein


```

Db 937 -FTKSENSNRTCSMPYAK---LEYK88N-----DSL-----NSVSSSDGY 973
Qy 320 PRRAKESKATREYERK-----QFP-EIRKORELOERMOSRGVQSGSLMSARSE 369
Db 974 GKRQCMKPSISYSYSDDESKEFCYSQGYPADLAHKHSANHMDDDGEJDTPTINSLKSD 1033
Qy 370 HEVEIITDGLSEOELEKOMQOLAVIPMLYDADQORIKFINNGIAMDPMKVYDROVM 429
Db 1034 -----EOLNSGRQ-----SPSQNERMARPKHIIIDEIKOSEFORQR 1069
Qy 430 NMWSEOE--KETFEKFMQHPKFNGLIASFLERKTVAECVLYYLTTKNENKSLVRSY 487
Db 1070 NQSTTPYVYESTDDKHLKQPHG-----QCEVSPYRSRANGSETRVGSNH 1119
Qy 488 RRRGKSQQQQQQQQQQQQQQQPMFRSSQEBKEKEKE-----AKKEEKEPVND 541
Db 1120 ---GINQVQSQSLQCEDDYEDDKPTNYSERYSEEOHEERPTNYSJKYNEBKHWOP 1176
Qy 542 KEDLKEKTD-----DTSGEDN--DEKAVASGKRKTANSGRRKGRITRMA 587
Db 1177 IDYSLKATDLPSSOKOSFSPSKSSGQSKTEHMSSESTSTPSSNAKQNLHPSSA 1236
Qy 588 NEANSEEAITPOQSALASMEINESSRWTEEBMETAKGLLEHGNMGAIRAMVGSKTVS 647
Db 1237 QSRSGQ---FOKATCKVSSINGTIGTYCEDTP-----ICFRCSLSLS 1281
Qy 648 QCKNPFYNYKRRQND--EIIQOHLKMEK--ERNARKKKKAPAAASEEAPPPVEDE 703
Db 1282 SAEBEIGCQTOEADSNATLQIAEIK-EKIGTRABDPVSEVPVSOH-----PRTKSS 1335
Qy 704 EMEASGVAGNE---EMVEEABALHASGNEVPGE-----CSGPATVNNSSDSTEPS 753
Db 1336 RLQSSLSSEBARHKAIVFSSGASPSKSGAQTPTSPREHYQETPLMFSCSTVSSIDS 1395
Qy 754 PHTEAKDTGONGPFPATLGADP---PPGPPTPRRTSRAPIEPTPASEATGATP 808
Db 1396 FESHSIASVQSEPCSGWVGIIISPSDLPSGQMPSPRSK-----TP 1439
Qy 809 PPAPPSAPPPVVPKEKEEBETAAPVEEGEOKPPRAEBLAVDTGKA--EEPVKSE 865
Db 1440 PP-PPQATQKREVPKMK-----APPAEKRESGPQAAVAAVQVPLPDAOTLLHF 1491
Qy 866 CTEBAEBC-----PAKGDAEABEATAEGALKAKEKGGSGRATTA---KSGAPQDS 915
Db 1492 ATESTPDDFSCSSSISALSLDEBPTQKQVLELRIMPVONDNGNTESEOPKESNENOK 1551
Qy 916 DSSATCSA--DEVDEABEGD---KNRLSPRPSTLPTGDPBRANASPOKPLDKQLKOR 969
Db 1552 EAEKTISEKULDDSDDDIEILEECIISAMPT-----KSRKAKKP---AQT 1597
Qy 970 AAALPPIQVTKNHPREDAAFTKAPAPAPPPQULQPSDAPQOPSS--PRGKSRSAP 1028
Db 1598 ASKLPFPAPARKPSQULPVYKLLPSQ-----NRLQPOKIVSFPQGDMPR--- 1640
Qy 1029 PADKAPAAABQKULGDPFCMTSGLPFPVPPREVIKASPHAPDPAFAFYAPRPHPLGL 1088
Db 1641 ---VYCVGEG-----TPINFTTA-----TSL 1657
Qy 1089 HDTAAPVLPAPRPPTISNPPLISSAGHPSELERQIGALSQMSVOLHVVYSEHAKAPVPV 1148
Db 1658 SDL-----TIESPP-----NELAAGEVNRGAAGSEBKKDITPT 1692
Qy 1149 TMLGLPMDPKLAPFSGVQKQBSLPRGOAPPSLGVF---TAQASVLKGTALASV 1203
Db 1693 -----BGRSTDEAQGGKTSSVTIPELBDNKAEBGDILAEICINSAM 1732
Qy 1204 PGG-----SITKGIPTRVPSDSAITR-----GSITGTADVLVYKGTIRI 1246
Db 1733 PKGSHKFRPVYKTIQVQOASASSAPYKNOLOGKKKPPSPVPIQONTETFRVYK- 1791
Qy 1247 IGEDSPSLDRGREDLSLPGKHVITYEGKKGHVLSYEGMSVTQCSKD---GRSSGPPHE 1303
Db 1792 -NADSKNNLNAERFVSDNK-----DSKKQNLKNNSKDFNDKLPNNEDRVRGSAFADSPHH 1845

```

```

Qy 1304 TAARKTYDMMEGVRGAISASIEGI-----MGRAIPREHSPHILKEQHIRGSI 1355
Db 1846 -----YTPLEGPPYCPSRNDSLSLDPDDDDVLSR---EABELKAKENKESEAKV 1894
Qy 1356 TQGIPIREVAQEDYLREAREKTLKEEGTPP-----PPSRDL-----TEAVYKQ 1400
Db 1895 TSHTELINSQOSANKQALAKQPIINRGQPKIILQKQSTFPSSSDIPRGAATEKQONF 1954
Qy 1401 ALGPLKLPRAHEGLVATVKEAGRSIHETPREBLNHT-----PELPLARPLKEGSI 1452
Db 1955 AIENTPVCFSHNSLSLSIDQENNNKKNEMPIKETEPDPSQGPSPKQASGVAPKSFV 2014
Qy 1453 QCTPLKYDTGASTG-----SKKHVRSILGSPGR 1482
Db 2015 EDTPVCFSRNSLSLSISIDEDDILQECISAMPKPKPSLKDNDENHSPRNNGIIGE 2074
Qy 1483 TFPVHPDLVWADARALERACYEESLKSPPCTASSGGSIRARGAPVIVPELKGPROSPLT 1542
Db 2075 -----DLTLDKDIQRPSEHGLS--PQSEMFDMKALQEGANSIVSSL----- 2115
Qy 1543 YEDHAPRAGHLP-----GSPVTMRBEPTPLQEGSLSSKASODRK 1584
Db 2116 ---HQAAAACLSQASDSDSLILSKGISLGSPPHL---TPQOEKXFTSNKG--- 2164
Qy 1585 LSTPRELAKSPHSTVBNHHPISPYEHLRGV--SGVDLYRSHIPLAFDPTSI PRGIL 1643
Db 2165 ---PRILKPEKSTL-----ETKKIESKGIKGGKVVKS----- 2197
Qy 1644 DAAAYVLPRLHLPNPTVPHLPYLLIRGYDPTALENRQTIINDYITSQQMHNHTATAM 1703
Db 2198 -----LITG-----KYRSNSELSCQW----- 2213
Qy 1704 AQRADMLRGSPRESSALNLYAABRGITIDLSQVPHLPVLPPTPGPATMDRLAYLPT 1763
Db 2214 -----KQPLQANPISISRG---RTMIIHIGV----- 2236
Qy 1764 APQFSSHSSPLSPGCPHTLTKPTTSSGERDRDRERDRERREKSLITSTTT--- 1820
Db 2237 ---RNSSTSPVSKKGP--LKTTPAKSPSEGO-----TATSPRG 2273
Qy 1821 ---VEHAPIMPRTQSSGSGSGSGGSSSRPASHAHQHSPIRPTODALQOR 1874
Db 2274 AKPSVKSLSLVAQ--TSQIGSSSKAPBSGSRSTBRAPQOLSRITQSPGNSI--- 2329
Qy 1875 PSVILNTEMKGITAVEBSKPTVLASTSTSPVRPATFPATHCPLGCTIDGYR--- 1930
Db 2330 ---SPGRNGI---SPPNKLSQLPRTSSPSTA-----STKSGSGKMSYTSPPGRQ 2373
Qy 1931 ---TLMERVLLPKKAPRVARPERPRADTGHAFLAKPPARSGLEP-----ASSPKGSE 1980
Db 2374 SQOULTKQGTGSKANSST---PRSESAGLQOMNNGNANKKVELSRMSSTSGSE 2428
Qy 1981 ---PRPLVPVSGHATITARTPA--KNLAPHNASDPDPAAPASDPHREKTQSKFPIOE 2035
Db 2429 SDRSERPLVNGOSTFIKEAPFTLARKLEBASLESLSPPSRAPSPRTSQOQT--PVLSPS 2487
Qy 2036 LELBSLGYHSSYSPEGEVPSVVSPELTHDKGLPKHLELDKSHLEGLRPRKQOPGVK 2095
Db 2448 LPMDSLSLTH--SSVQAGGRKLPPLNISPTIENDGRPARHNDIARSHSPSRL---PIN 2542
Qy 2096 LGG--EAAHLPHLAPLPSQSPSSPLQOTAGVGNQHVUPLAQHSIVITQDTRHHPQ 2153
Db 2543 RSGTWKREHSHK-----SSSLRPVSTMRRTGSSSLSASSESSKAKASBDEKH--- 2591
Qy 2154 QLASAPLAPLXFPGASCPLVLDLRRPSDLVLPDHDGAPARGSPHSEGRKSPENKTS 2213
Db 2592 -----VASISGK-----QSKENGVSAKGMTRKIKENEFSTNBS 2627
Qy 2214 ---VLGGEGDIE-----PVSPPGOMTEPRGSRSAVLPVLYLRDGBQTERPSMGSKSPONT 2265
Db 2628 QTVSSGATNGAESKTLIYQMAPAVSKTEDVWVRITEDCPI-----NNPRSGRSPTGNT 2679

```



```

QY 1809 EREKSLTST-----TVEHAP-----IWRPTE-----QSS-----GSSGSGGCGG 1846
Db 2162 TSLASVSTSAPEGTSASVAPKLSLSPDVSQSTKTFDATESSTVOASSTSGTSEV 2221
QY 1847 SSSRASH-----SHAHQSP1-SPR-----TODAAQGRPVLYHNTGMGIIITA 1889
Db 2222 STSEESHVTKLITSSNPSVSPVTSPTVEBSTEQPTSTTPSGOSLTPMNSSEV 2281
QY 1890 VEPSPKTVLSTSTSPVPAATFP-----PATRCPLG-----GLDGYPTL 1932
Db 2282 LTTSEPHVL-SSLSLSDVQSSTTPNNLESSTVEPEKTSVSVLNSEBPTTEA--PTT 2338
QY 1933 MEPLVP-----KEAPVAPRERPAOTGHAFLAKPPARSGLEPAS-----SIS 1976
Db 2339 LSPDLITSTNNLSQSTVSTSETRSEISSENS--EKPTAPRPLVTSVTHVASSSPDYPT 2396
QY 1977 KSEPPRL-----VPVSGHATTARTPAKLAHPHNASPPDPAPASADHREKTOGK 2029
Db 2397 ESSEPDLTGSSSTENIPEASKOTISSTP-----TPD-----TTTSEPTSTMS 2443
QY 2030 PSCIOELRLSLGHSYSYSPGVEPVSPSP-----SLTHDKGLPKHLEELDKHLEGE 2085
Db 2444 P-----DLSTTSNVLSSSSTTPES--SSKSPVSSSTEGISVVTSEFSKVESTISSVLEBD 2498
QY 2086 LRPKQGPVYKLGSEAAHLPHLRPLPSQSSSPILQTA--PGVKGQRVTLAQHT---S 2140
Db 2499 LTKTPTSP1-----LEETTTASETSEPLETDSLTVSVRIHE-LTTSSENVPKES 2546
QY 2141 EVITQDYTHHPQOLASPLPAPLYSPGASCPVLDLRPPSDLYLPPDHG--APANGSPH 2199
Db 2547 ESTTSSSESKPQEPAGILITSVVPTSSVSLITSELEALITNSNPFQGRPTITSPK 2606
QY 2200 -----SREGKRSPEPNKTSVLGGGEGDIEPVSPGEMTEPHGHSANVPLIYRDGEOT 2252
Db 2607 SLVKTSTSPETVSSSEPSSEKTKRTVSTVSTTPEETTT---SESLI--LTAAPSKPT 2661
QY 2253 EPRMSKSPGNTSQPAPFSPKLTESNSAMVSKQEIINKKLTNHRNREYNI-----S 2307
Db 2662 ESTTESSEAP-----TTPAKTSEKPSNVSTSRKSTE-NVEISTSGSGSLESSTMSSTG 2716
QY 2308 QPCTEIFNPMATGTGIMTFRSOAVOEHAFT-----NMGLAIIIRKALMKYDQWEE 2359
Db 2717 EPERT--NAPAYVSSSEAS--STTLENSSTSPSPTSSEASVKLSLFPESTSEAVTGS 2771
QY 2360 SPP-----LSANAFNPLNASASLPAPMPTAADRSHTLTPGCGGKAKVSGPSSR--- 2412
Db 2772 RAPAEITMSSESHREISTVSSSEPEIPLSTVSPNVVTA-----SSIPSEPTL 2822
QY 2413 ----KAKSPAPGLASGDRPVSVSSEBDCNRTPLTNKRWEDRPSASGTFP--PYNPL 2467
Db 2823 SSVTSSSTPRVRLITGTPDLIVSVTVPSHGKRRKNT-----ASVYSNGTSP1 2872
QY 2468 IMRLQAGVMASPPP-----PGLPAGSGELAP 2494
Db 2873 ILPSESLTTPQPPPTTTTAKPATTSKGRP 2903

```

RESULT 54

RBHUP

adenomatous polyposis coli protein - human

N/Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jul-2000

C/Accession: A37261; B39658; A4928; A49319; I54271

R/Kinzler, K.W.; Nishio, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith

chui, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A/Title: Identification of APC locus genes from chromosome 5q21.

A/Reference number: A37261; MUID:91335210; PMID:1651562

A/Accession: A37261

A/Molecule type: mRNA

A/Residues: 1-2843 <RIN>

A/Cross-references: GB:M74088; NID:g182396; PIDN:AAA03586.1; PID:g182397

```

R.Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grod
arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Adderahlm, H.; Cohen, D.; Le
Cell 66, 601-613, 1991
A/Title: Identification of deletion mutations and three new genes at the familial polypc
A/Reference number: A39658; MUID:91330307; PMID:1678319
A/Accession: B39658
A/Molecule type: DNA
A/Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P'
A/Cross-references: GB:M73548; NID:g190163; PIDN:AAA6054.1; PID:g190164
R.Miki, Y.; Nishio, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst
Cancer Res. 52, 643-645, 1992
A/Title: Disruption of the APC gene by a retrotransposon insertion of LI sequence in a c
A/Reference number: A44928; MUID:92119623; PMID:1310068
A/Accession: A44928
A/Molecule type: DNA
A/Residues: 1506-1525 <MTK>
A/Cross-references: GB:S78214; NID:g243541; PIDN:AB21145.1; PID:g243542
A/Note: sequence extracted from NCBI backbone (NCBI:78214, NCBI:78218)
R.Spiro, L.; Olschansky, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber
Cell 75, 951-957, 1993
A/Title: Allele of the APC gene: an attenuated form of familial polyposis.
A/Reference number: A49319; MUID:94073973; PMID:8252630
A/Accession: A49319
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 'G', 143-171, 'P', 173-179 <SPI>
A/Cross-references: GB:S67787; NID:g461061; PIDN:AA013997.1; PID:g461697
R.Lambert, S.; Ballhausen, W.G.
Hum. Genet. 90, 650-652, 1993
A/Title: Identification of an alternative 5' untranslated region of the adenomatous poly
A/Reference number: I54271; MUID:93186137; PMID:8383094
A/Accession: I54271
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-4 <LM>
A/Cross-references: GB:S56365; NID:g266243; PIDN:AA014918.1; PID:g4262770
C/Genetics:
A/Genes: GDB:APC
A/Cross-references: GDB:119682; OMIM:175100
A/Map position: 5q21-5q22
A/Note: mutations of this gene can result in familial adenomatous polyposis or sporadic
C/Suprafamily: adenomatous polyposis coli protein
C/Keywords: cancer; familial adenomatous polyposis; tumor suppressor
F.1-730/Domain: leucine-rich <NTD>
F.7-72/Region: coil #stratus predicted
F.185-227/Region: coil #stratus predicted
F.731-283/Domain: serine-rich <CTD>
F.1131-1156/Region: acidic
F.1558-1577/Region: acidic
F.1866-1893/Region: highly charged

```

Query Match 2.7%; Score 350.5; DB 1; Length 2843;

Best Local Similarity 17.8%; Pred. No. 6.5e-05;

Matches 473; Conservative 357; Mismatches 1007; Indels 821; Gaps 109;

```

QY 50 ASHLSFGSLTOP--QRRPSLSEFO-----PGRSGQELHLPESHSY 91
Db 740 ANIMSPGSSLPALHVRKQKALFAELDAQHLSFTFNIDNLSPKASHRSKQKSLYGDY 799
QY 92 LPELKSSEMEFESKRPRLLELPRLRSPLLAQCPGSGEDL---KDRSL----- 141
Db 800 VFDYNNRHDNRNDNFNTGMVTLSPLNTVTVSSSSSSSGSLDSRSSEKDRSLERBGIG 859
QY 142 TGLKLEVPSPSPPHPTDPELELVPRLSKEELLQNDMDRVDRITWVEQOISKLKKQOOLE 201
Db 860 LGNYHATEN-----PGSSSKGL-----QSTTAQAIKTV-----ME 892
QY 202 EBAKPEPEKPEVSPPIESKRSVLQIYDENR--KAAEAHRLIEGIPQVEYPLVNO 259
Db 893 EVGAHHTSQEDRSSGSTB-----LHCVTDERNALFRSSAAH-----THSNYTN- 936
QY 260 PSTROYENIKINQMRKLLVLPFRKRHAKQKQKCCQYVDQMEALEKKVERIENN 319

```

Db 1813 GGGGVCFGTSAATTAATATGSGFCQAGSGSNTGVSFGQAAS-----TGGIVFGQGS 1866
QY 2428 PSVS-SVHSEDDCNR 2441
Db 1867 SSSSGSVFGSGNTGR 1881

RESULT 53

T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34513

R:Ravello, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Releases: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Intons: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;

3504/1

Query Match

Best Local Similarity

Matches

476; Conservative

288; Mismatches

879; Indels

848; Gaps

117;

QY 550 TDDTSGEDNDEKAVASKGRKTAASQGRKGRITRSMANSEBAPPOQSAELASML 609

Db 715 TEEDEGEDEDMEEGSGSMSTTINGTG-----ITGSPREGTIRVRIT-----TL 759

QY 610 NESSRWTEEMETA-KKLLLEHGRWMSAIAWVGSKTYSOCNFFYFNKKNONDELIOQ 668

Db 760 GE-----DGEPEATKPGISAPDKTGE-----GSKTESD-----GE 790

QY 669 HKLKMEKE-----RNARRKKKKAAPAAASEAAPVVEDEMEASGVSGNEEMV 718

Db 791 EKLTVEDKGAQSGSSATSGKSEATSGSSSSA-----KSGTSAASGSG----- 840

QY 719 EEAALHAAGNEVPEGSGSPATVNNSSDTESIPSPHTEAAKDTG----- 763

Db 841 -----ASSSSGSGVSGSSSVTBSGSGFTSSSGSVGEATGSGTVDGSESGKPKSSTE 896

QY 764 -----QNGPKPRATLGDGPPPPPPPTPPRRTSRAPIT----- 795

Db 897 EKLPTTKNGEKSPIS-GSD--TTGKSSSEETTSRPRIGSDSLTEGSGSGEWFTGSKQH 953

QY 796 -----PTPASEATGAPTPPPAPPSAP-----PPV 821

Db 954 FESGSKSVTSGKGT-OSGAGSGSGKPKVPKGPAPETITDGERSSSTSGDKSGKPA 1012

QY 822 -----VPR-----EKEKEETAA-----PVEEGEGOKPPAAE-- 849

Db 1013 DKSDKNVPTKGDKNPDITTDGEDSTSETSGEGQPKKSKQPPGDKGSEVKKPTSEVD 1072

QY 850 -----ELAVDTGKAEEVSECTEBAEG-----PAKGDAAEAETAGAL--KAEKEG 898

Db 1073 GPGNLSGKGSNVPLKP--TDLPEEGSGILTTSSGSGNSTFEHGTKLERLPKPKEDSS 1130

QY 899 GS-----GRATTAASGAPQSDS-----SATCSADEVD--EAEGGDNRLIS 939

Db 1131 ETPQLGLEISAKKKEPEDGSKVEGLIWESTTPGSTTIDSDVGLFISGDLTK-AT 1189

QY 940 PRPSLL--TTGDPFRANASPOKPLDKOLKORAAAIPIPIOTVYKVEPPRED----- 988

Db 1190 KKPVEIEGSGTGDSEITATTR-----DVSKSTKKPRVEVDGDNGB 1231

QY 989 --AAPTKAPAPPPOQLQPE-SDAPQPGSPGSKRSRPPAPPADKFAAQAQKLP-G 1044

Db 1232 TSGVDGKRTTPAPTPSSAESTSRIPPTSEASPEGGSGEAGVESPDSGSSSTGADBG 1291

QY 1045 DPCCWTGSLPPPPPREVIKASPHAPPSAFSVAPPGHPLPLGHDTPAPVLPPT--I 1102

Db 1292 VSP--TSSATAPPEVPTTSASTPDAVESSG-----ISTSKPTAEPLATTAPSTEV 1340

QY 1103 SNP-----PPLISSAHPVLEBOIGAIQGSVQ--LHVPSHAKAPVGVPT 1149

Db 1341 TSPGSGTEESTLPTTEGSGE-----STTSAPVPEBAPVLPQNRNEKEDEPTKDT 1391

QY 1150 MGLPLPMDPKLAPFSGYKQQLSPRQAG-----PPESLGV-- 1186

Db 1392 FALPTTTGADQANDSSVENTKCTSSDGCALCERRTVCRCGPRGEGAPPKGSCVDV 1451

QY 1187 -----PTA-----QEASVLRGTALG----- 1201

Db 1452 DECATGDHNCBSARCONVGVGACFPTGRKXADGSCODIDECHNSTCCGANAKCV 1511

QY 1202 SVPG-----GSITKG--IPSTRVPSDAITRGSITHTPADVL----- 1238

Db 1512 NKPGTYSCECENGLGQGYQCVPTTKPCDST--QSSKSHCSBSNMSCEVDVTVDGSVEC 1568

QY 1239 -----YKGT-----ITRIIGDSPSRLDGRDLSLPKHVITYEGKKGHVLSYEG-GMS 1285

Db 1569 KECMGYKSKKVCEDINECVAEKAPCSLNNANVNM--NGTFSCGCKG--TRGDSFM 1622

QY 1286 VTQCKEDGRSSSGPPE-----TAAPKRTYMEGRVRA 1321

Db 1633 CTDINECDERHPCHPHACNLTBGSFKCECHSGEGDGIKCTNPLBESCSGVKFCGRV 1682

QY 1332 --ISSAST-----EGLMGRALPPPHSPHHLKEQHITGST-----TGTI 1359

Db 1683 DHVCLSVRIYNGSLSSVCECEPFRFEKESNSCVIDECEBSRNCDPASAVCVNTG- 1741

QY 1360 PRSY-VEAQEDLYLREAKLKRGEPTPPPPSRDLTEAVTQA-----LGP--L 1405

Db 1742 --STRCEBAGY-----EGEG--GVCITIDEDCRMGACDSAMCINRGSGCC 1786

QY 1406 KLPKPAHEGLVAT--VKEAGRSIHEIPREELRHTPELPLAPRPLKEGSIITGTPLKYDTG 1462

Db 1787 KOMAGYTGDATCTKIEBERPSDCTACTDEMSRLCELEKQCTYDEEVPQ----- 1837

QY 1463 ASTGSKKHVRSLLGSPGRTPPVPHPLVDNADAPALERAC-----YESLSKSRGT 1514

Db 1838 -----CGACLPGHHPINGTCOSLQISGLCAQRKDNCKNAECIDIHPDS 1880

QY 1515 --ASSSGSIRAG-----APVVP 1531

Db 1881 HFCSPDGFIDGMCICDDVDECNAGMCDENTYCENTIGSFNCVCLBGPKNVDEKCVVD 1940

QY 1532 ELGKPROGSLTYEDHGAFFAGHLPRGSPVTYRREPTPLQESLSSSKASQDRKLTSTPRE 1591

Db 1941 EKKQPNREKIIDENS-----SSNSGQEKPTTK--GIVSTASTSSESTAAE-- 1988

QY 1592 IAKSPHSTVPBHHPHISPYVHLIRGVSCVDLYNSHPIPLADPTS--IPRGIPLDAAAY 1649

Db 1989 -----HVTTSISSTSTK-----DMTSSKPEVNTVMSBS-- 2018

QY 1650 YLPHLAPNPTVPHLYPPYLRYGVBDTAALENROTIINDYITSOQMHNHTA----- 1700

Db 2019 -----PEVSTSSKSTTJASSETTVSSTPSESSSEAPLNTSSP 2054

QY 1701 --TAAQRADMLRGLSPRE--SSIALNYAAGPGIIDLQOVPHLPVLPPTPGPATAM 1755

Db 2055 ATTTEVITESSVSKSTTPRESSSEITVLSKSPSEVLESS-----VKSSPSTPTS 2106

QY 1756 DRLAYLPTAPQPFSSR--HSSSPUSPGGPTHL-----TKPRTTSSSERBRDRDRDR 1808

Db 2107 QSVT--STVPEITSKTVLSSEAPVPTSTSPTEVHTSSETKPBLASAS--TTGDNTSTPTS 2161

Db 2199 K-----NE-----MFRKEILRR-----LDLILIELL-G 2220
Qy 2353 KYDQWEEPPPLSANAFLNPLNASLPPAMPITPAADGRSDHTLTPGGGGA--KYSGRPS 2410
Db 2221 AEDBEDOKRDLK-----QIP-----TSEEDTDSKADSMGAECSAFRRILSRSS 2264
Qy 2411 SRKAKSPAPGLASGDRPPSSVSHSEG 2437
Db 2265 TWGNNGSGSPS-ASGTTSPSTSSISSG 2290

RESULT 52
S26058
Probable transforming protein (can) - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S26058
R: von Lindern, M.; Fornerod, M.; van Baal, S.; Jaegle, M.; de Wit, T.; Buijse, A.; Grosveld, M.; Cell. Biol. 12, 1687-1697, 1992
A: Title: The translocation (6;9), associated with a specific subtype of acute myeloid leukaemia.
A: Reference number: S26058; MUID: 92195315; PMID: 1549122
A: Accession: S26058
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-2090 <VON>
A: Cross-references: EMBL:X64228; NID: g29652; PIDD: CAA45535.1; PID: g29653

Query Match 2.7%; Score 355; DB 2; Length 2090;
Best Local Similarity 20.6%; Pred. No. 3.2e-05;
Matches 362; Conservative 177; Mismatches 656; Indels 560; Gaps 79;

Qy 993 KPAPAP-----PPQNLQPSD-APQPGSSPRGKR 1024
Db 381 KTLPPAPVLMILSTGVLCPFYMINQPGVKSILKTPERLSLEGERQPKSPGSTPTTPPS 440
Qy 1025 SPAP-PADEKFAFAEQKLPP-----GDP 1047
Db 441 SQAPQKLDASAAAAPASLPPSSPADLAFSLIPAGCAPTVSFGSSLSKSAATVTGEP 500
Qy 1048 CWTSGLPFPVPPREVIAKSPHAPDPSPAFSAPRGH---PLPLGLHDTAPVLP----- 1097
Db 501 SYSSG-----SDSSKAP-GPGPSTFVPPSKASLAPTP-----AASPAVASASPS 547
Qy 1098 -----RPTTISNPPPLIS-----SANGPSVLERQIGA-----ISQMSVQ 1132
Db 548 FGSSGFKPTLESTPYPSVASAPNIAKMSFPPSTSAVKVNLSEKFTAAATSTPVSSQSA 607
Qy 1133 LHVPSSEHAK-APVGVVTMGLPLPMDPKTLAPSGVKQEQQLSPRQAGPPESLGVPTAG 1191
Db 608 PMSPPSSASKPAAAGPLSHPTPLSAPSSVPLKSSVLP---SPSGSAGSSSPVPSMTQ 664
Qy 1192 ASVLNCTALGSVPGSITKGIDISTRYPSDSAITVRSITHGTPADVLYKGTITRIIGDS 1251
Db 665 KSP-RITPPAAKPGSQAKSL-QPAVAEKQHQMKSS-----DEVMAG-----IGEE- 709
Qy 1252 PRLDGRGRDLSLPGKHVITYEGKKGHVLSYEGMSVTQCKEDG---RSSSGPHEHTAAR 1307
Db 710 -----IAHFOKELEBELKARTSKACQVGTSEEMKMLRTESDILHTFLLE 753
Qy 1308 -KRTYDMEGRVGRATISSASIEGLMGRALPREHSHHLEKQHNIIGSTTQGI-PSYSYE 1365
Db 754 IKETTESLHGDIS-SIKTLLRGFAG--VEEARQENRNDSGYLHLKYRPLDPKS--E 808
Qy 1366 AOBVLYRAEAKLKRREGTPPPPPSRDLTEAYKTYQALGPLK-----LKPAREGLVATVKE 1420
Db 809 AAGQETRLKHQYKF-----AVQDVNDVDLDMDOQLBQKKQRHLVBERETLFTLAN 863
Qy 1421 AGRSIHEIFREBELRHTPELPLAPRLKE---GSITGTPCLKYDTGASTGSKKHQDVRSIL 1477
Db 864 -NREITNQQRKRLNHLVDSIQQLRLYKQTSIMELSSAVP-----SQSSIHSPDSDESIC 917

Qy 1478 GSPGRTPPPVPLD---VMAADARALERACYESELKSR--PGTASSGSGSIARCAPVYBE 1532
Db 918 NALLKTTIESHTKSLPKYPAKLSPMKQQLNRLFLAKRTTPVPRSTAPASLSAFL----- 973
Qy 1533 LKPRQSPULYEDHGAIPAGHLPRGSPLYTMEPTRLQEGSLSSSKASODRKLTTSPRE- 1591
Db 974 -----SORVYEDL-----DEVSTSSVSQSLSESDARTSCKDEAVVOAPRRA 1016
Qy 1592 -IAKSPHSTVPEHHHPISPY--EHLR-----GVSGVDLYRSH---IPLAPPTSIPRGI 1641
Db 1017 PVRKTP-SIQPSLPH-AAPPAKSHLVHGSSGCVNGTVAISASKIIPQADSTML----- 1070
Qy 1642 PLDAAAAYLPRHLAPNPPTLYPPLRYGVPDPALENQ-----TIIN 1687
Db 1071 -----AKTVGHGAPSPSHPISAFOQL-----AAALLRQMAQAPAVNTLTSTLKN 1118
Qy 1688 --DYTSQOMHNHTTAAQADMLRGLSPRESSLALNYAAGPRGIDLSQVPLVLP 1745
Db 1119 VPQVNVQELKNNPAPTP-----STAMGSSVPYSTAKTPH-FVLTP 1157
Qy 1746 -----PIPGTPATAM-----DLAVLPTAPQPFSSRHSS 1775
Db 1158 VANQAKQGLSLNLSKPGPTPASQSLSSGKASGTAKETVAVTSPSASQFSKPFSS 1217
Qy 1776 PLSPGCPHTLTKPTTSSSERERDRDREREREKSLITSTTVEHAPIWRPGTEQSS 1835
Db 1218 PSGTGFRGIIPTTSS-----NFTAAQGA---TPSTKESS 1250
Qy 1836 GSSGSSGGGGSSSPASHSHAHQHSPISTQDALQORPSVLNHTMKGITTA--VESS 1893
Db 1251 QPDASFSSGGG--SKP-SYEALPESP--PS-----GITSASNTTGG 1286
Qy 1894 KPTVLRSSTSSPVPAAPFPATPCPLGGLDGVYPTLMEVLLPK----- 1940
Db 1287 EP-----AASSSRPVAPSGALSTTSKL-----ETPPSLGGLLPPSSLAGTGLSFGSL 1337
Qy 1941 ---EAPRVARPERPADT-----GHAFLAKPPARGLEP--ASSPSKGE 1980
Db 1338 RVGQADDSIKPTNKASSTLSGTPTKTSGVPSGFNFAPVYLGNKTEPPTVSSATTTSV 1397
Qy 1981 PRLVPPVSGHATARTPAKNL-----APHNASDPRA--PP 2015
Db 1398 APPAATSTSTAVFGLPVTSAAGSGVLSFGTSLAGKTSFSFGSQQNTSTVPPAPRP 1457
Qy 2016 ASASPDRKTKQSKPSTIGELRLSLGYHSSYSPGVEPVPSGSPSTLTHDKGLPKHLE 2075
Db 1458 TTAATP-----LPTSFPTLSFGSL-----LSATTPSLPMAG--RSTE 1494
Qy 2076 ELDKSHLEGLRPKQGPVYKLGEEAHL--PHLRPLPSQSSP-----LIQTAPG 2125
Db 1495 EATSSAL-----PEKRGDEVSASAAASLLEBQGSQQLPQAPQTSVYKKEPVLAQPAVS 1549
Qy 2126 VNGHORVTLAQHISEVITQDTRHHPQQLSAPLAPLYSFG----- 2168
Db 1550 NSGTAASTSLVALSEATPATGVDPARTBETVPPASFSFVGQTAVTAAALISSAPVAV 1609
Qy 2169 --ASCPVDLRAPPSLLYLRPPDHGAPARSGSHSECKKSPRPNTKTSVLGGEDGIEPVS 2226
Db 1610 ETTSSTPIAS--STTSIYAPGSAEAAAFGTVSSGSVFAQPPAASSSSAFLQLTNTNTAT 1666
Qy 2227 PREGMTEPGHSAVAYPLLYRDGEOT-----EBSRMGSKSPGNTSQPPAFSKL 2275
Db 1667 ABSATPFGQVAVASTAPSLF--GQQTGASTAAATTPQVSSSGFSPAPAGTTAPGVFGGT 1724
Qy 2276 T-----EBSNAMYKSKQOELINKLNTNRNEPVYNIQPG--TEINMPAITGT 2322
Db 1725 TFGQASVFQGSASSASV-----FSFGQPGSSVPAGQPA--S 1761
Qy 2323 GMTYRSQAVQEHASINMGLALIRKALMKXYDQWEEPPPLANFNPLNASS----- 2375
Db 1762 STPTSTSGSVFAASST-----SSSSSFSFGSSSNTTGGLFGGSANAPFGQSPGF 1812
Qy 2376 -----SLPAMPITPAADGRSDHTLTPSGGGGKAKVSGRPSRKAQAPGLASGDRP 2427

QY 514 RSQOEKDEKEKEAEKEEKEPVENKED---LKEKTDJDSGDNDKEAVASKRK 570
 Db TRSVEHTHTNOEDENNASGSDSSDSEEGSSSSNEDSDQONVDEDEDDVSEKKR 468
 QY 571 TANSOGRRK--GRITRSMAN-----EANSSEALITPOQSAELASMLANESSRMTSEEMETA 623
 Db HEPEBGKASSPONGHRDSNGDKHEDSSERSQOSTSH-----HETSHPREDSLAY 522
 QY 624 KKGLEHGRMWSALARMVGSKTVSQCKNFYNYKKRONLDELLOQHLKMEKERNARRKK 683
 Db 523 Q-----SRSPSL-----NYQOSPGYEFLSEKIKOE-----FS 552
 QY 684 KKAAPAAAEBAAPFVVEDEEM-----EAGVSGNSEEMVBEAALHAS--- 727
 Db 553 PTTSSASSDLELDMEMPDNPLTMLERMHMRPFIDVSPVNRIDEIVELQOKARASYEK 612
 QY 728 --GNEVPRGEGSPATVN-----NSDPTESIPSPH 755
 Db 613 FTGRFPFK--CNDEVLISIQKIVHEPRDYTYENPCSELEVRIDMKLSTLADLDR 670
 QY 756 TEAAKDTQONGPKPRATGADGPRPGPTPRRTSRAPLEPT--PASEATGAPTPPPAP 813
 Db 671 ATDSKEIGRDQOPAGRTSGR-----PSLDESRTNRLSPDSTHHAELAQSHSLCIGPM 724
 QY 814 SPSAPRPVVPKKEKEEETAAAPRVVEGEBQKPAABELAVDTGKAEBPVKSECTEABEG 873
 Db 725 TPSTP-----PFTSQPLLVNTTH----- 742
 QY 874 PAKGDAAEATAGALKAKEKGGSGRATTAKSSGAPQDSDSATCSADEVDEAEGD 933
 Db 743 -----LPGTSQBSTSGITTPRS----- 760
 QY 934 KNRLSPRPSTLTPTGDPANASPOKPLDKQKORAAIPRIQVTKYHEPPREDAAPTK 993
 Db 761 -----SQPPLMSPVSRHNSMSTGRPASIQTLRHQSVWFP-----DVSIFP----- 803
 QY 994 PAPPAPRPQNIQPPSDAPQOPGSSPRGSKSRPAPADKEAPAAAOQLPGPPCMTSG 1053
 Db 804 --PPIP-----THDMWAPR--GTPPSRSSETWPLSPPGTPIQNL-----LTM 847
 QY 1054 PPPVPPREVIKASPPAPDPASFAFVAPRPHPLPLGLHDTAPV-----LPRPTISNP 1105
 Db 848 PIVPPPHLIAATSTGTHSVSSASHSTPRST-----SGTPVHCPSNKSQTPTPKSR 901
 QY 1106 PPLIS-----SAKHPSVLERQIGALISQMSVOLHVPSSEHAKAPVPTWGLPLPMDK 1159
 Db 902 PEKVQIRHDTISKSGPSNAINALQARSQM-----TSG-----DBK 937
 QY 1160 KLAPPSGVKOE---OLSPRGOGAPRESIG---VPTAQEASVLRGALSGVPGGSITTKGIP 1213
 Db 938 KSAPSTPVVNRDAGSLVAQIMSNOP--NLGLRLPRIEKKSSALONIQNHQPHSANSTP 996
 QY 1214 STRVSDSAITVRGSIHTGTPTADVLYKGTITRIIGDS-----PSRLDRGREDSLPKG 1266
 Db 997 ST--PSTS-----TH-----QAMFKDKEKRRKKKEKERESEBARBEMKKTKEER 1042
 QY 1267 HVIYEGKKGHVLSYEGMSVTQCSKE--DGRSSSGPRHETAPAKTYDMMEGVGAISGA 1325
 Db 1043 NKRKEMERAKRLEDERQERKERKEKEDERKKEKEREKKAKEKELKKKGHRKGDSDBS 1102
 QY 1326 STEGLMGRALIPPERBSPHHLKEQ--HHT-----RSGITOGIP--RSYAEAOEDYLKRA 1375
 Db 1103 DSDSNDDELIDVRSKTKEMTOEKDHQIALLLLSKGLIENLKSRRRSDRRADSEFKIQ 1162
 QY 1376 KLLKR-----G-----TPPPPSRDLTZYAYKTQALCP 1404
 Db 1163 KSQGRVLLESDEGKDGDKGNSNGESESSEKADLPPPAFBS--LSEADQ----- 1216
 QY 1405 LK-LKPAHEGLVATV-----KEAGRSIHE-----I 1428
 Db 1217 LKVLKEREKGEJLTSSDEDEHDNAG--EIHQQLTEDRENRKROKSLTAVSSDEGERKNV 1275

QY 1429 PREBLRHTPELPLAPR---LKEGSIITQGTPLKYDTGASTGSK--KADVRLISGPGRT 1483
 Db 1276 PKMRWRDSEDAAKHPMSAKDQOKRKLERNRSEDESSKNAKADFPDI----- 1328
 QY 1484 PPPVPHLDVMDARALERACYEESLSKSPGTASSGGSIAR-----G 1525
 Db 1329 -----PHEDVDEEBTE-----DGRSRRQSTSSITSNVATAKERKESGKPLRIVPEPTG 1379
 QY 1566 APVIVPELAKRQSPLYTEDGAPRAGHLPRGSPVTMEPEPTPL-----QEGSLSSXASQ 1581
 Db 1380 TPLSPKILSPK-----HL-----SPKTSITSRKSSISDHENLISROKNR 1421
 QY 1582 DRKLTSTPRELAKSPHSTVPEHNPRISEYEHLLRGVGVLYRSHIPLAFDPTPIPGI 1641
 Db 1422 TTSSTSTATSSKHALSIFP--KPLSPVTATSSVSID-----DP-SIRDEF 1467
 QY 1642 PUDAAAAYULPRHIANPPT--YPHLYPYLLRGYPDOTALNROTIINDYITTSQGMHNT 1699
 Db 1468 SMNSAAD-----SPMSTGRPMVLTGAKAKAFNSTPKKETDQAVQSIFDEE----- 1515
 QY 1700 ATMAAQADMLRGUSPRESSIALNYAGPRGIIIDSQVPHLPVLVPTPTPATMDKLA 1759
 Db 1516 ADEFPQYPDF--GISTEKEYS-----GKDPH--NIKPT----- 1545
 QY 1760 YLPTAPOPSSRHSPLSPGSPYTLT--KPTTSSSEBERD--RERDRER----- 1810
 Db 1546 -----EPLNNGHTDILFSPSSAHASEKOSTKESDDHEBDSLVYMEKEVMEVIAOE 1599
 QY 1811 -----EKSILTSTTV-----EHAPIWRPTEOSGSGSGSGGSGSSRPA 1852
 Db 1600 VHVPEPSPMEBEVLTETSPVPKKEPPIKMESSBEGTTPRDIISNBSQDTG--A 1652
 QY 1853 SHSHAHQH-----SPI--SPRTODALQO--RPSVLNHTMKGKITTAVERSPETYLKSTS 1902
 Db 1653 VNNHLENHDAVQPIIQOAPSOHQVAPSRPAAPDSOQNPVLVSQOSQSPSSSQO 1712
 QY 1903 -----TSSPVPAAT--PPATHCPLGLDGVPTL-----MEVLLP 1939
 Db 1713 SDMAONLLSSKDVIDLAKHKNRBEALAOATRGDCSSIPOHLLHAQNGQNMTPENIQ 1772
 QY 1940 -----KEAPRYAR----- 1947
 Db 1773 LKAAPFAQOEENEAQOMQAKKQOTINKDRIKEQERVKMYEENRKYEDREKORKE 1832
 QY 1948 PERPR--ADTGHAFIAPKAPASGLEPASP--PSG-----SEPRPL--VPPVSGHA 1992
 Db 1833 BEKORLAATAATWATQAKAEBALKQOEVRHGFQHVLSMTPPARSLYEQFPGLSSYI 1892
 QY 1993 TIARTPAKNLAPH-----HASPDPAPASDPHREKTSKSPSIOELER 2039
 Db 1893 NRDSIGATNGVILHPTQOSTORPSSSTASTSSNPKAPLQPSASVNOQT--IDPAIEIRIVQ 1951
 QY 2040 SLGYHGSYSPEGVPEPVSFVS---PSLTHDKGLPKHLEBLDK----- 2079
 Db 1952 RMFYKPLKMSAEBAATWAAVASSDENPRATSVVDLAAMLQOLQAAQAAQOVVUTTA 2011
 QY 2080 -----SHUEGLRPKQOPVKLGEBAAHLPHLRPLRESQSSSPYLQTAPEVKHQVY 2133
 Db 2012 STPNPLSMILETLTASTASLANLATGA-----LNPILSMIALTSS--LNOSSPYUQIARVL 2064
 QY 2134 -----TLAON--ISEVITQDVTYTRHPOOLASPLPAPLYSPFGASCPLYDLRRPSDYL 2185
 Db 2065 LTMNGOMLATYQISELLA--TMNQETLMAALLA-----RNLPRAM 2105
 QY 2186 PPPDH--GAPARGS-----PHSEGKRSPEPKNTSVLGGEDGIEPVSPREGMTEPCH 2236
 Db 2106 PQONOQPMQAPAGGAFIPTVLPHM--SLKRNADQUS--VGVSVD-----R 2147
 QY 2237 SRSAYVPLLRYDGEQ-----TESRMSGSKSPGTSOPRAPFSLCTLTSNANMYKSKQOEINK 2292
 Db 2148 KKSCLPLAMIGQGOQPPPOQPMQAVAPAPSPSP-----RKSMEFNLPREMKE 2198
 QY 2293 KLTNHRNREPEYNISQPTGEIFNMPAITGTGLMTYRSQAVQVHASTNMGLAETIRKALMG 2352

OY		306	MEALEKVERI-----ENNPRRAKESVREYYKOPPHIRKORELOEORWOSVGQRG	358
Dd		585	LKSLEKENKAIKCGGSETSRGRGRQOLPKPTEKETOTLQ-SLOKENELMKSLEGULE	643
OY		359	SGLSMSAARSHEHVESIIDGLSEOEENLEKOMROLAVIPMLYDADQRIKFIMNGLMAD	418
Dd		644	TPL-PCGTENOELVSSIOENLESITALEKENOPLASPEV---GBEALRPLTKEN--QE	697
OY		419	PMKYVKROQVMNMWSQEKETFR--EKPWOHPRONFGLIASFLERKTVAECULYYVLTKCN	476
Dd		698	PLRSILED-----ENKEAFSLSEKENOEPL-----LKLTLEEEOISIVRPLET	737
OY		477	ENYKSL-----VARSVRRCKSQOOOQQOOOQQOOOQMPMRSSQOEKEKEKEAE	530
Dd		738	ENHSKLSLEBODEPTLTLEKTQOKRSSIGEDODWTLP---PEKVDEPLKSLD	791
OY		531	KEEEKPEVNDEDLLEKTDUTDSGEUNDEKEAVAASKRTYANSQGRRRKGRITRSMANEA	590
Dd		792	OEIAR-LENNOEFILSKIES-----VEAKSILETILS-----LKSAGE	834
OY		591	NSEBAITPOQSABLASMELNSSWTBEEM-----ETAKKG-----LLEH	630
Dd		835	NLETLSBPETOAPL-----WTPELIKSGNSKRKSGSRITTVGCSSPPRIQT	884
OY		631	GNNMGAILRMVGSKTVSOC-----KNFYFNKKRONDELIIQOHKLXMKERNARRK	683
Dd		885	CRGESGITIEIGSMERGFEISRGVDKESCQNLEEBNLKGXEOSTL-----RSLBERG	939
OY		684	KTAAPAASBEAAPPVVEDEME-----ASGVGNENEVEEHALMASGN--VPBR	735
Dd		940	QBLPOSADVORWEDVTXDOELAESPPGMAGVENKDEALNTLEBDQFTGKEEVBOGE	999
OY		736	CSGPATVNNSSUTESISPSP--HTFAADTOQONPKPKPATL-GADGPBPPTPPRRTS	790
Dd		1000	LMA-----TEEWMFGEGHPNPNPKRGRLVGGASVKGGAEGIQ-----	1039
OY		791	RAPIETPASEATGARPTPPAPPSAPRPVVYKKEKEEETNAAPRVEEBEOKPPAAE	850
Dd		1040	----DBEQSQOVGTGLOAPQGLPEALBPVEDVAPGDQDASPEWILSSE--PAMGE	1092
OY		851	LAV-----DTNGK-----ABEPYKSCBT-----EASEESPAGKDAAEAAT	886
Dd		1093	SAAAGAPEGLQGCVGLIDPCHILTREEWNEPPLRESLEAKVOGLEBPR-KOLEEA---	1147
OY		887	AEGALXAEXKE-CGSGRATTAKSSGAPODSDSATCSADEVDAAE-----GDKRNLLSP	940
Dd		1148	--GGLOTERTSELPGKSRDPWEPPREGRESEAEARGAEAPPAETLGHGTSD-----AP	1200
OY		941	RBSLLTPGTDDPPRANASPCKEPLDLKQLKRAAIPLOY--IKYHEPPREDAAPTKPAPPA	998
Dd		1201	SP---WPLSEEEED-----VPVULVSPEPTTYTLIEDA-----	1232
OY		999	PPEPONULOPESSAPOPGSSPRGKSRPAPPADEKAFAEAOKLGPDPCWTSGLPPVP	1058
Dd		1233	----PBLQOABESQASMGVGYGRAEAGVSESQIELDS-----GEIP--EGL-----	1274
OY		1059	PREVIKASPHAAPPSAFSTAPRCHPLPLGHTTAPVULP-RPTINSNPPLISSAKH---	1114
Dd		1275	QEEGESRESEDELDGETLPDSTPLGFYLRS--PTSFRMTLESRGHLPLAKTGKBCWD	1331
OY		1115	PSVL-----EROIGAISQMSVOLHVPSY--HAKAPVGPVTMGLPLPMDPKLT	1161
Dd		1332	PAYLASBLEPSEPKEGEBEGECCORDDLSBEFPDLGTEAFP-----LGVPGEV	1383
OY		1162	A-PSPGVKGOQLSP-----RGQAGPSPSGC--	1185
Dd		1384	AEPICGYQPLLDPAAWDRDGESDGFADREESGEBGEDOEGBRPGAARMGPSGSVSL	1443
OY		1186	--VPTQOASVLAGTILG-SVP-----GGSITGIPSTRVPSSD--AITYGSIHTCTPA	1235
Dd		1444	QAUSSSGRGEPLSDSVSVPMDDSLRKANA-VGAPKTLTFRESODNASPSSBEEESPV	1502

QY 1236 DVLVYGTITRLIIGEDSPSLDRGRSDLSFKGHVIT-----YEGKKGHVLSIEGG-MSV 1286
 Db 1503 SLEREDKV-----FGPLTIPSGMEDAGADITIGVGNCPNLIEGSGQH--NGVWNG 1553
 QY 1287 TQCSKEDG-----RSSGSPHETAPKRTYDMMEGRVGRALSSASIEGIMGRAIP 1336
 Db 1554 LEQSSSESGARNALVSEGDGSPPOEEGSAIKR-----SSAG----- 1590
 QY 1337 PERHSPHLKEQHIRGSIYTGIPRSYVEAQE 1368
 Db 1591 ----APVHLGQGGQFLKFTQREGDRESWSSGED 1618

 RESULT 51
 T20531
 Hypothetical protein F07A11.6a - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T20531, T27776
 R:Palmer, S.
 submitted to the EMBL Data Library, October 1995
 A:Reference number: T20531
 A:Accession: T20531
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 12526 <MTL>
 A:Cross-references: EMBL:Z66511; PIDN:CAB54210.1; GSPDB:GN00020; CESP:F07A11.6a
 A:Experimental source: clone F07A11
 R:Gajdstry, S.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: Z20417
 A:Accession: T27776
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2526 <MT2>
 A:Cross-references: EMBL:Z69904; PIDN:CAB54501.1; GSPDB:GN00020; CESP:F07A11.6a
 A:Experimental source: clone ZK20
 C:Genetics:
 A:Gene: CESP:F07A11.6a
 A:Map position: 2
 A:Intons: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1

 Query Match 2.7%; Score 356.5; DB 2; Length 2526;
 Best Local Similarity 18.0%; Pred. No. 3.5e-05;
 Matches 490; Conservative 338; Mismatches 972; Indels 927; Gaps 123;

 QY 155 HTDELELVPRLLKELIQNMDRVDRREITWEEQISIKLKKQOOLEE-----AAKPE 209
 Db 47 HGNBEIYLVSTMDARGAL-----PARSTRKPEEDSIEYKISAYIBE 87
 QY 210 PEKEVSPPIESKRSLSVQIITYDENRKKRAEAAHRLLEGIGPQVELPLVNPQSDTRYHN 269
 Db 88 PTQN-SPIP-EFPESESTACVVE-----IQSGTPE-----RDLFEL 122
 QY 270 IKIQMAWEKLLIVFKRNHARKKOKFCQRYDLMELAEKVERIENNPRRAKESKV 329
 Db 123 VK-RHKSRSQVPIDIQLESTTEPGHKKARVHYRLDTDSLKADKSLILGR-----PKF 175
 QY 330 REYY-----EKQPEIRK-----QRELQERMQSRVQORG 359
 Db 176 RYVYPTSEQKHPCQHPSTSYAIPKLKGDHLKASCVHYVPHLDHRSPOHYRRFESYQ 235
 QY 360 GLSMSAASHHVSSIIDGLSEQENLEKQMRGLAVIPMLY-----DADQ 404
 Db 236 VIDVMVSNDKAAVAVOFTNIDDAQKALQDTNIPKPMHSYQSRPSHRHIIIFLPIECTN 295
 QY 405 QRIKPI--NMGLMADPMKVVYKQVMNMWMSQEKEETFEKEMQ--HPKNFGLIASFLER 460
 Db 296 EIMILIRISLSRIYIDICVDMWDRSAVITLDMIEPANLLIKRWKLIVGRNNFG-----EH 349
 QY 461 KTVAE-C--VLYYLTIKNENYKSLVRSYRRRGSGQQQ--QQQQQQQQQQQQQPM 513
 Db 350 KYAVDFCGDRFNLVYINRKENIE--VAARSSSEFTSKSENDQSSSPSSSDRQNLHPDQ 408

```

Db 777 TSGSTQTPGSSVSTSAIILTSTQGSVSTNPGSTVTPRSTVSSGSSSTVTVGSTAS 836
Qy 1104 -NPPPLISSAKHPVLERQIGAIQGMVQLHVPSSEHAKAPVGTWGLPLPMDP--- 1158
Db 837 TSGSVASSPAPSTSQNPNPSTSGSSMITQSPYPSQSTSPVEBST--TPSPGSPGTTL 894
Qy 1159 -----KKLAPSGVKEQJLSPRQAPPPESLQVPTNOEASVLRGTALG 1201
Db 895 TSTSPSPSQSTTIGTQGSTPBGISTTSEMTSQSTQTPGSTQSTVTPSTSDSTSSG 954
Qy 1202 SVPGGSIKYG---IPSTRVPSDAITYRGSIHTGT-----PADVLYKG----- 1241
Db 955 STTVVGTSGSSSPITPSTSQNTNPSTSGSSMSTQTPSSQSTSPVEBSTSGATSSGSP 1014
Qy 1242 --TTRIGEDSPSLDRGRDS-----LPKHVIEGKKGHVL-----SYEGG- 1283
Db 1015 GTTLTISIPSPSPSTIGSSQSTSPVVTISQGSTETPGSTGTVTPSTVSGSASGS 1074
Qy 1284 ---MSVTCQSKEDGSSSGP-PHEIAPKRYTDMMEGVGALISASIEGLMGRAPPER 1339
Db 1075 TATMGSTIASSTSGSSSTSPNPSQSTSPS-----TSGATSSPGSSGTTLTISP-- 1123
Qy 1340 HSPHLKQEHIRGSIQGIPRSVEAQEDYLREAKLKRGTGPPPPSRDLTEAYKT 1399
Db 1124 -SP---SQSSTIGS-SQG-----STSPVGTSGDMTSSQST 1155
Qy 1400 QALGFLKLPKPAHEGLVATVKEAG---RSIHEIPREELRHPPLPLAPPL----- 1446
Db 1156 QI-----PGSTGTVTPQSTGSGSTGSGTSGEITQSGSTQTPRSSLSLSPALSTQOSV 1208
Qy 1447 ---KEGS-ITQCTPLKYDT-----GASITG-----SKKDVSLLISP--- 1480
Db 1209 STNPSGTVTPQSTVVRSGSTSGSTVTTGTEBSSSTSGSSSATSLSSSPVSTQSPNPS 1268
Qy 1481 --GRTPPVHPLDVADARALERACYERSLKSPCTASSGSGSIRGAPVIVPELQKPRQ 1538
Db 1269 TSGSSTPFPNP-----SQSTSPVSTTGTGMTSHGSTQTPSTIGSTVYT 1311
Qy 1539 SPLTYEDHGAPPAHLPRGSPVYTKRPTPRLOEGSLSSSKASQDRKLTSTPREIAKSPHS 1598
Db 1312 QPST-----VSGSNSGSGTVTI-----GSSEASTSGSSPK--TSPESI---S 1348
Qy 1559 TVPEHHHPHPISEYHLLRG-----VSGVDLYRSHPLAFDPHSIRGIPLDA----- 1645
Db 1349 PVPSTSPPTSTTPASTSGSTISDVSVST-TSLAPLS---SSLSTVPSSTQSSSTSE 1404
Qy 1646 ---AAATYLPRLHAPNPYPLPYPYLIRGYDPAALENRQTIINDVYTSQOMHNTAT 1701
Db 1405 GSKKASSSPVPSQSTSTPNP-----TGSTES-STLSSSTISGSGTGH---T 1446
Qy 1702 AMAQADMLRGLSPRESSIALNYAAGPRGIIIDLSQVPHLPVLVLPPTGTPATA-MDRLAY 1760
Db 1447 TMSK-----ASSGSTSPSTNSQTG---STV-----TMSGSSSTGVSSTSA 1483
Qy 1761 LPTAQPPSSRHS-----SSPLSPGSPTHLTKTPTTSSSRERDRDRDRDRERERS 1813
Db 1484 SSTQOMSTSQSSAGSTVASTAPASSTAPASSTGTMSS----- 1524
Qy 1814 ILTSTTVYHAPIMPTGTEQSSGSSGSGGSSSRPASHAHQHPISPRTDALQO 1873
Db 1525 --TSGTGSTISSESTTASASQSTGYVTWSSSGSTSGVSTSSASTQOMSTQSSSAG 1582
Qy 1874 RPSVLHNTGMKGIITAVEBSKFTVLRSSTSSPVRPAATFPATHCPLGLDGYPTLM 1933
Db 1583 STVASTAGL--VSTSTVPSSTGTWGSTSGT-----VSGSTI----- 1617
Qy 1934 BEVLLPKAPRYARERPRADTGHAFIAPPARSGL-----EPASSPGSEPRPLV 1985
Db 1618 -----SESTTASAS--SQSTGYVTWSSSTSGVSTSSASTQOMSTQSSG--- 1662
Qy 1986 PVSCHATTIARTPAKMLAPNHASPDPPAPASAPDHREKTSQKPSIOELRLSLGYNG 2045
Db 1663 ---SAGSTVA-----SSTTGIVSTSTVPSSTQOMSTSGTGV 1697

```

```

Qy 2046 SSYSPEGEVPEVPSVSPSLTHDKGLPKHLEELDKSHLEGELRPQPGPVKLGGEAHLPH 2105
Db 1698 STIS--ESTTASASQT-----GSTVWSSSTS--- 1725
Qy 2106 LRPLPESQSPSSPLLQTPAGVKHQRVVTLAQHTISEVITDQYTHHPQOLGAPLPAPLYS 2165
Db 1726 -GVSTSSASQPGQPMSTSQSSAGSTVSS----- 1753
Qy 2166 PPGASCPLYDLRPPSDIYLPPPHG-APANGSPHSBGCKSPPEBNKTSVIGGEDGIEP 2224
Db 1754 --STAPASASTASSTGTMSTSSGTVGSTMOSSTASTSHGTGTVTLGSSSTSSNQ 1811
Qy 2225 VSPPEGMTEPHSSAVVPLLYRQGEQEPERMOSKSPGNTSQPPAPFSKLTESNAMYK 2284
Db 1812 MSTQSSVSGSTVASTAGLVSTVPSSTGTMSTSGTGVG-----STISESTTASA 1865
Qy 2285 SKQEIINKLNTNHRNEPEVNIISQPTIEIFMMPAITGTGLMTYRSQAVQEHASTMGLA 2344
Db 1866 S-----SQTGSTV-TMGSSSTGVSSTSSASTQPMSTSQG--- 1900
Qy 2345 IIRKALMGKYDQWESPELSANAFNPASASLPAMPITLADGRSDHTLTSPGG--GK 2402
Db 1901 -----SSAGSTVASTAGLVSTVPSSTGTMSTSGTGVSTIS 1941
Qy 2403 AKVSGRPSRKAQAPGLAGDRPPSVSVHSEGDCKRRPILTRVWEDRPS-AGSTP 2461
Db 1942 SSTASTSQSTGYVTLTGSTGTMSPSPSLSQ-----ITTPSPSQSTESTQTSLPSSSP 1997
Qy 2462 PPYNPLIMRLQGVNAS 2478
Db 1998 SPSTHVSASSSGTMMSS 2014

```

```

RESULT 50
S21424
nestin - human
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: S21424
C:Jaccard: 0.7; McKay, R.D.G.; Zimmerman, L.B.; Lendahl, U.
R:Jaccard: 0.7; McKay, R.D.G.; Zimmerman, L.B.; Lendahl, U.
submitted to the EMBL Data Library, May 1992
A:Description: Characterization of the human nestin gene reveals a close evolutionary re
A:Reference number: S21424
A:Accession: S21424
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1618 <DAH>
A:Cross-references: EMBL:X65964; NID:g35018; PIDN:CAA46780.1; PID:g35019

```

```

Query Match 2.7%; Score 358; DB 2; Length 1618;
Best Local Similarity 21.1%; Pred. No. 1.9e-05;
Matches 323; Conservative 196; Mismatches 511; Indels 502; Gaps 77;

Qy 97 KSEMEFISKAPRL--ELLP--DPLRPSPL---LATQPA--GSEDTLTKRSLTKLEP 147
Db 329 KLELPFPPTPEGRKRGSLPLPILSPSLPPLPATLETVPAPLAKQBFLOARTPLASTP 388
Qy 148 VSPSPRPTDLELVLPRLSKELLQNMMDRVDRBITWVEQOISKL-----KKKQOOLE 201
Db 369 I-PPTPQAPSP-----AVDAEIRADQADPLSLQTOGGRKQAPPLR 428
Qy 202 EEA--AKP-----PEPEK-----VSP-----PRIESK----- 223
Db 429 AEARVALIPASVLPGEPEGQROEASTSQSPEDHASLAPPLSPDSSLEAKDSEGGSRV 488
Qy 224 -----RSIVQIYDE---NRKKAEAAH---RILEG 248
Db 489 PSICGEBEGQIWGLVEKETAIEGVVSSLOQETIEBEDIRAKELQDQGVPLPEKTKLSL 548
Qy 249 GPQVE--LPLYNQSDTRQYHENIKINQAMRKLLILFKRRNHARKQMKQKFCQRYDOL 305
Db 549 GEIQESIKLLENQSHETLE-RENOECPRSLIEDL-----ET 584

```



```
QY      1546 HGAPRAGHLPRGSPRYTMEBPPRLBOEGLSSSKSAQDKLITSTREIAK--SPHSIVPEBH 1604
Db      646 ATPP--GKKP-GSPPSFRTGTTPGYRG-----TSPPAGPGTFKPS--PTVG 687
QY      1605 PHFISPYEHLIRGVSVNLYRSHTPLAFDPDPSIPRGILDDAAAAAYLLRHNAENPTYHNL 1664
Db      688 FGFL-----PAGFGELP-----SLPPPAANAS 711
QY      1665 YPPYLIRGYPDTALENQOTIINDYTISOQHNTATMAQRADMLRGLSPRESSALANY 1724
Db      712 GPP-----LSAQIKOEPAEEYET----- 730
QY      1725 AAGPRGIDLSQVPHLPVLVPPTPCTPATMDRLAYLPTAROPFS--RHSSSPSPGAPT 1783
Db      731 ---PESPYPAPASPPP---PKVDVPEHAQSANFNHLDGRFSCARSIDLTYFTPLEGS 784
QY      1784 HLTK-----PTTSSERDRDRERDRERBKSLITSTTVE--HADIW 1827
Db      785 KLAKRADLVKEVRREAORAREEKEREKEREREKERERLESEVRVKLOABGRAIVE 844
QY      1828 RPTGESSGGSSSGCGGSSSRPASHSNAHQNSISPRTDALQORSVLIHNGMKII 1887
Db      845 CP-----SLGPVPH----RPFEFGS--AAVTVPYL-GEPTPALR 878
QY      1888 TAVESEKTVLRSTJSTSSPVBPAAATFPPTHCPIG---GTLDGYUPTIMEVILLPKCAP 1943
Db      879 TLSEYARBEHVMSPCGRNHMF-----VYPLCAVDPGLLGVNVVALYSSDPARAERE 928
QY      1944 RVAPBERRADTGNAFLAKPPARGSLGEPASS--PSKGSRRPLVPPVSHAIATTAANTL 2002
Db      929 REAEERDLRKLKGFGEVXP--SELBRHGVPGGILD--FRHHGGLAQOGPGGLHP 982
QY      2003 APHNASDP-----PAPPASADPPHEKTQSKPFISIOELTELRSLGHSSSYSPEGVE 2054
Db      983 FFFPHSLGLEPERRLALAAGPALRD-----MSYAE-RILAERQHAERYAALGND 1031
QY      2055 FVSFVSSSSLTHDKGLPKHLEBELDKSHLEGSLRKQEPQVYKLGGEAANLPHLRPLPESQP 2114
Db      1032 PLARIQMINTV-----PNNHOH--SHISHHLHQODAHNAASAVH-PLIDPL----- 1077
QY      2115 SSSPLLQAPAVNKGHQWVVTLAQ-----HISEVITODYTRNH-----POOLSAP 2158
Db      1078 ASGSHLTRIIRPAG-----TLPNPLPHPLHENVTL-----RHQLFAARYDDLPAASLASAP 1127
QY      2159 LPAPLYSFFGASCFLVLDLRPPSD---LYLPPPDHGA--PARGSPHSEGKSREP 2209
Db      1128 MSA-AHOLQAMHASOAEIORTALECOOWTAAHHNPYSVLPFAQGDYVSHLKEKSDKP 1183
```

```

RESULT 47
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.muscl.edu/gsc/c_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-7738 <STO>
A:Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:G3881547; GSPDB:GN00020; CESP:F07A11.6
A:Gene: F07A11.6
A:Map position: 2

```

Query Match	2.8%;	Score 368.5;	DB 2;	Length 2738;
Best Local Similarity	18.1%;	Pred. No. 1.4e-05;		
Matches 436;	Conservative 211;	Mismatches 885;	Indels 803;	Gaps 94

Qy	44	HHSDYASHLSBGSLIQORRRPSSL	-----EFQONE---	RSQELHLPESHSTYLP	93
Dd	635	HEPRDYUYEENPCSELLEVRIDRWKLS	STADLDLDFRADTDSKEIGDOPAGCTSGRPSLD		694
Qy	94	ELGKSEMEFFISKR	-----RLLELPDPLLRSPPLATGQAPAGSEDLTGDRSLTGKLEPV		148
Dd	635	ESRTRRLSFDSTHHPALAQSRSHSLCIG	PTMTPTSTFPPLSQPL	-----LVNTTHLPCTSQS	750
Qy	149	SPSPPHTDPELELVPPRLSKEELI	QNMDRVREITTWVQOISKLKKQOQOLEEAAKPP		208
Dd	751	T--GGGITPTPSSOPPLMSPVSRHNS	MSGTGRPASI	-----QTLRHOSVMP	796
Qy	209	EPEKVPSPPLESKRSLVQIIYBE	-----NRKXEAHAHRLLEGQPVQELPLVNO		259
Dd	797	-PDVSI PPPPLPPH	-----DMMAPKGTPEBRSSSTW	-----VPLRPP	835
Qy	260	PSDTRQYHENIKINOAMRKULIYFKRNNHARKQWKQFCOR	YDQJMEALEKKVRIENN		319
Dd	836	PFGT	-----PIGNLLTMPILVPPH	-----LIATSTGTHSVSS	869
Qy	320	-----PRRAAKSKYR	-----EYEEKOPPELRKQRELOERNQSR	-----VGQSGSLMSMA--AR	367
Dd	870	AHSTPRHSISGTPHACBPSNKS	TSPPTPKSR--PEKQIHRDITISKGSPNAIHALQAR		927
Qy	368	SEHEVSEIIDLSEQENLEKQMRQALAV	PPMLYDADQORIKFINNN	-----GLMADPMKY	423
Dd	928	SQSMTS	-----GDPKKSABSTPYVRDAGSLVQIMNQNGLGRKPRLEK		974
Qy	424	KDROVNMWMSOEKETPEFEKFMQHPKNGFLIASFLERTVA	BCVLYYLTKKNYKSLV		483
Dd	975	KSSALQNIQNHQPHSNANSTPTSTSTHQA	MFQDKEKE	-----RKKEKEKEER	1022
Qy	484	RRSVRRGKSQQQQQQQQQQQQQQQQQMPRSSOEKDEKEKEK	-EAEKEEKPEVEND		542
Dd	1026	EERARREMKRKEYTEERKKRKEMEKAKGLJEDROQRKKEKKEK	ERBERKEKEKVKAKKEK		1089
Qy	543	EDLLEK	-----TDDTSGENDKEKAAVAASKGRKTANSQGRKGRITRSMANEANSEAI		596
Dd	1086	EKLKKKGHRKGDSDSDSDNDLDDLVRS	-----TKEMQE		1122
Qy	597	TPOOSAEILASHELWSSRWTEEBETAKGLLEHGRM	SALARMVGSKTIVSQCKNFYNY		656
Dd	1125	--EKDHOALAL	-----LSKGIIE	-----NL	1143
Qy	657	KKRONLDEILQOHLKMEKERNARKKKKAAPAAASEAA	PPVVEDEMEASGVSGNEEBE		716
Dd	1144	KSRRSRDGRANDSFEKQOQKSGQRVLIES	-----SDDGKGDDGNGS--		1187
Qy	717	MVEBAELHMSGNEVPRGECGPAITVNN	SPTESI PSHPTAAKDTGONGPPKPAITGLAD		776
Dd	1188	-----SNGEE	-----SDSEK	-----AD	1199
Qy	777	GPPEPPTPPRTSRAPLEPTPASEATGAPPPRAPPSAP	PPVPVPEKEEBEETAAAP		836
Dd	1200	LPPP	-----PAPPLSLSBADQLKYLKEREKELTTSSDD		1233
Qy	837	VEEGEOKPPAAEELAVDTGAKEBPV	SECTEAEGBGAKG	-----KDAEAAEA	885
Dd	1235	-EDNNDAGEIHOQRLTEDRENRKKQKSILTAV	SDEQGRKNVPKRMRRDDSEDAAKKPG		1293
Qy	886	-TARGALKAKEKGGSGRATTAKS	-SGAPQSDSATSADVEABEGDKNRLLSPRS		943
Dd	1294	WSAADQOKQRKQKLEHRRSSDEDSKNAKRFDI	PHEVDSDEEBTEGBSRSKQSTST		1355
Qy	944	LITPTGDPANASPOKPLDKQLQORAAL	PIPIVTVKHEBRDDAAPTKAPRAPAPPPQ		1003
Dd	1354	ISNVTAKEREKSGSKTPLRIYVPEBTG	PLPSKLPSKHLSPKSTSTKXS	-----	1400
Qy	1004	NLOPESAPQOQSSPPGKSR	-----SPAPRADYEAPAAEQKLPGRDP	PCWTSGLPFPV	1058
Dd	1406	-----SISDBHNLTISPQRNMKTSS	TATATSSGH--EALSTIPEK	-----PLS	1446

453 HSTARRRGSRRRSRGSRADRRGS-----HRSRREKRRR- 491
QY 1262 SLPKGHVIEGKKGHVLSYEGMSTQCSKEDGRSSGPPHETAPKTYTMEGRVGR 1321
Db 492 -----RRSASPP-----A 500
QY 1322 ISSASIEGLMGRAPPERHSPHHLKEQHIRGSIQGIQPRSYEAQEDYLREAKTLKRE 1381
Db 501 ASSSS-----SSRREH-----RKRREG-----GKKKKSRSAEGRS 535
QY 1382 G-----TPPPPPSRDLTEAYKTOALGPKLKPANBGLVATYKAGRSIHEIPRELRHTP 1437
Db 536 GDLEKLPAVPSPSGSDRSRRGAVP-----SIQDLTDHDLFAI- 575
QY 1438 ELPLAPRLKESITQITPLKYDTGASTGSKKDVRLISPGKTPPPVPL-DYMDA 1496
Db 576 -----KRTIVGRPDKTEPRAP-----SPA---PAVSPKGVLYDS 608
QY 1497 RAL---ERACYEESLKSRRPQTASSSGSARGA---FVIVPELGKPR---OSPLTYEHHG 1547
Db 609 EGLADERGAKGDKORRRSGAASSSSSRKSRKALDGRDRDRSSKPRTPKD-S 667
QY 1548 APPAGHLPRGSPVTKREPTPLQESLSSSKASQDKLTSTPRE---IAKSPHSTVPEH 1604
Db 668 APGSALPKAP-----RSGSSSSSSSSCRKXKLOSKVALLREGVSTTPAKD 717
QY 1605 PHPISPYHLLRGVSGVDLYR---SHIP-LAFDPTSIPIRGIPLDAAAAYLPRHIAPEPT 1660
Db 718 SSSSG-----LGSIGVFSRDRBSRSPFLKPDERSPAEGV-----KVAPEST 759
QY 1661 YPHLPVYLIGVPTALENNQTIIINDYITISQCHNHMTATMAQRADMLRSLRBSGL 1720
Db 760 KPKTKVAKAKAKAKAKGTGK-----TKPSKTRKVVRSYGSGSTAGSGPSLKKSKAD 812
QY 1721 ALNTAAGRGIIIDLSQVPHLPVLPPTGPTATAMDRLAYLETPAPQSSRHSSPLSPG 1780
Db 813 SCSSQASAKGTEETWSGE-----ERTKAPSTPPKVAPEPPALTPDSQVDSCKTP- 866
QY 1781 GPTHLTKPTTSS-----SERERDRERDREREKSLITSTTVENHAIWKPCTEBS 1835
Db 867 DVSPFLAEASBDTVGVVGAEEEEESEEEEOQPAITATSTAAAPSTAP----- 920
QY 1836 GSSSGSGGSGSSSRPASHAHQHSPIPRQDLOQRPVLAHNTGMKGIITAVEPSKP 1895
Db 921 -SASSTAGDSGEDPAA--SOLPT-----LPPMP 950
QY 1896 TYLRSTSSPVPRPATPPATPCPLGTLGTVPTIMEPVLLPKRAPPVAPERRA-- 1953
Db 951 NWL-----PAGVDCITTSGL-----ALTALLFKMEANLA--SRAXQE 987
QY 1954 ---DTGHAFL-AKPPARSGLEPASP-SKSEPRP---LVP---FVSGHATTIARTPAKNL 2002
Db 988 LIQDNTQILRHKKPSTLGVTPAPVTPSGLPAPASSYLLPGLSLIGGSGSTPPPT-GL 1046
QY 2003 APHAASPPRAPPAA-----SPHREKTSKRFPSIQELELRSLGN-GSYSPREGEVPS 2057
Db 1047 VP--ASDRKREGSSSEGRGDTKYLKLTQERAAVEVLSIKFYQKKDITKEEYKOL 1104
QY 2058 PVSSTSLTHDKG---LPKHLLELDKSHLE-----GELR-PRQGPVKLG 2097
Db 1105 KRAYVKICHSGKGEINPVKVSUVAIVYVRKYGKPKGDPGPPRPKEPBPDPKG 1164
QY 2098 GEAAHLPHL 2106
Db 1165 GGLPLPLPL 1173

RESULT 46
G01763
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G01763

R:Margolis, R.L.
submitted to the EMBL Data Library, March 1995
A:Reference number: G08343
A:Accession: G01763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <MAR>
A:Cross-references: EMBL:U23851; NID:g915325; PID:g915326
C:Genetics:
A:Gene: GDB:DRPLA; B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p-12p

Query Match
Best local similarity 21.3%; Pred. No. 4.2e-06; Length 1184;
Matches 331; Conservative 136; Mismatches 605; Indels 485; Gaps 75;

QY 762 TGONGPKPATLGADGPPGPPPTPPRTSRAPIDPTPASEATGAPTPPAPSPSAPPV 821
Db 3 TRQKKDSMSMSGKKEKAPGREELRGRASPGVGSTSSDG----- 45
QY 822 VPKEKEEETAAPVBESEQKPPAABELAVTGAAEPVKSCTEBAEGPAKGDAA 881
Db 46 --KAESKQTAKKARVEASTPK-----VNKQGRSEHISESE--SEETNAPKKTTEE 94
QY 882 AAEATAGALKAKEKGGSGGANTAKSGAPQDSATSGADAV-----DEAG 931
Db 95 LPPQSPEDL-----DSLGRSLNDGSDPRDIDQNRSTSPSISYFSGVENDSDSSG 149
QY 932 GDKRLSPRPSLTPPTDPPRANASPKQLDLKOLKQRAAIPPIQVTKVNEPREDAAP 991
Db 150 LSGPARRYHPPPLPPEPPQPPDSTPRQ-----EASFPHPSVTPPGYAPMEPTSR 203
QY 992 TKPAPP-APPPOULOPE-----SDAPQPG-----SSPGRKSPRAPDAKFA 1036
Db 204 MFOAPRGAPRPHPOLYPGCTGVLGSPMGCKGGAASVSGPBGCKONPPPTPISVS 263
QY 1037 AEAQKLPDPPCWNSGILPFPVPRREVIKAS--PHAPDSAFSVAAPRGHPLGLHDTAR 1094
Db 264 SGA-----SGAPPTPPTTPVGGGULPSAPRANPHTNPLP----- 301
QY 1095 VLPPPTISNPPLISSAKHPSVLERQIGAISQGSVOLHVPYSEHAKAPVPTMGPL 1154
Db 302 -----PPRALRPLMNASASBPGLGA---QPLPGHLP-SPHA--MGQIGG--L 341
QY 1155 PMDPKK--LAPFSGVKQQLSPRGQAPRPSLGVPVTAQASVLRGTLASVPGGSTITG 1211
Db 342 PPGEPKPTLAP-----SPHSILPPASSAPADPMRFYSSSS--SSAAASSSSSSSSSA 395
QY 1212 IPSTRVPSDAITTRGSIHTGTPADVLYKGTITRIIGDSRSRLDRGKEDSLPKGHVYE 1271
Db 396 SP--FPASQALP---SYPHSFP--TSLSVSNQPKY--TOPLSQAV-- 436
QY 1272 GKKGHVLSYEGMSTVQCSKEDGRSSGPPHETAPKTYTMEGRVGRASISASIE-- 1328
Db 437 -----WSQGP-----PPPPY-----GRLLANSANHGCP 460
QY 1329 ---GLMGRAPPERHSPHHLKEQHIRGSIQGIQPRSYEAQEDYLREAKTLKREGT 1383
Db 461 FPPSTGAQSTAHPP--STHN-----HH-----HQOQOQOQOQOQOQHNGN 500
QY 1384 PPPPPSRDLTEAYKTOALGPKLKPANBGLVATYKAGRSIHEIPRE-----ELRHTP 1437
Db 501 SGPPPP-----GAPPHPLEGSSHNHAPVAMSSLSGLRLPY 537
QY 1438 ELPL-APRLKESITQ---GTPLYKYDTGASTGSK-----KHDVRL-IGSPGRTFP 1485
Db 538 PGPALPPPHSGVSTISQAGPNGPPVSSSSSSSTISQGSYCSHSPGCGQGA-YPFP 596
QY 1486 PVHPLDVNADARALERACYESLKSRRPQTASSSGSARGAPVIVPELGKPROSFLTYED 1545
Db 597 PV-FVTVTSSA-----TLSTVIATVASSPAGYKTASBPBPYGKAPSGAAYKT 645

```

Db      785 EKGIAGKPGILGTGA-----KGEAGYPPGPG----- 811
Qy      1280 YEGGMSVTCCKEDGSSSGPPHETAPKRTYDMWEGVRAISSASINGLWG-RAIPE 1338
Db      812 ----LPGKBPBGSTTGPPEPGPG-----LKGKDG-IPGAGLPGLSQKLP-- 857
Qy      1339 RSHPHLKEQHHIRGSIITQGIIPRSVYEAQEDYLRRKAKLLKREGTPPPPSDSLTEAYK 1398
Db      858 -----GVPGQ-----KSGILPGLAGAPGF----- 878
Qy      1399 TQALGPLKLPAGEGLVATVKEAGSIHETPRELHNTPELPLAP-PLKESGIIQGT- 1456
Db      879 ----GAKGEPGLPG-IG-----KSGPGPGQPGAPGPGQKGD--EGLPG 918
Qy      1457 -----LKYDTGASTTSGKKHIVRSIIGSPGRTFFPVHP-----LDVMAADRALR 1500
Db      919 LFGVSGMKDGT-----LPGVGLAGPGPGQPGPGQKQPGPPGAGAKG-- 963
Qy      1501 RACYESLSKRPQTASSSGSIARCAPVIVPELKGPRQSLTYEDHAPFAGHL- 1555
Db      964 ----EAGLPGLPGARPGQKQKQGLAGLPGIPKMGAGIP-----GAGQDGLPGLPGV 1012
Qy      1556 ----RGSPTWREPTPRLQESGLSSSKASQDRKLTSTPREIANSPHSTVEHHHPIS 1609
Db      1013 KCDRGFNGLPCKEKGEGPAPARDKQ-----EPGL 1042
Qy      1610 PYEHLLRGVSGVDLYRSHIPLAFDPTISRGIPLDAAAYLPLRHLPAPTYPHLYPVL 1669
Db      1043 PGQPGIRGQG-----PPGL- 1058
Qy      1670 IRGYEDTALENRQTIINDYITSQQMHNATATAMQADMLKGLSPRESSLALNYACPR 1729
Db      1059 ----GLGGLKDEG-----QPGYAGPLMGKQ--LGLPCKPBRPG--APGK 1099
Qy      1730 GIIDLSQVPHLVL-----VPTPGTPATAMDLAYLPAPPPSSRRSSPLSPG--GP 1782
Db      1100 G---LDGAPGFPGLKGEAGLPAGAPLP--GODGLPGLPG-----QKSGEGFPQPGPLVGP 1149
Qy      1783 THLTPTTSSSERENDRERDRERERKSLITSTTVEHAPIMRPGREGSSGSSG 1842
Db      1150 PGLP-----GKMGAPGIRGKGDGLPG 1172
Qy      1843 ---GGGSSSRPASHSHA-----HSHSPISPTODALQORPSVLNHTMGKGIITAVEPSK 1894
Db      1173 LFGEGGLDGLPGQKAGAPRGAPGAPGVPYKSGAGAPGF-----GLKG-----EPGL 1221
Qy      1895 PVLVASTSTSSVPRPATPPRATNCPGLGTLGVYPTLMEVLLPKRADRVARPERP-- 1951
Db      1222 PG-LGQPGPRGMKGEAGLPGA--FGRDGLPGL--FGMKGEAGLP-----GLPGQPGKS 1270
Qy      1952 -RADTGHAFLAKPRASGL-----EB-----ASSPSKSGSEPRPLVPVSGHATTAR 1996
Db      1271 ITGPKGNAGLPGLPGKDGGLPGLRGLKGBEPKPGVAGAAIKGEPGLPGIPGAKGEBGSLG 1330
Qy      1997 TPAKNLPANHASPPDPAPASADPHREKTOSKPFESIQELFELSLGYHSSSYSPGVEVY 2056
Db      1331 IPEGK--GNDGIPGKRPAGLPGP-----GKMGESGLPQCGRA 1368
Qy      2057 SPVSSPSLTTHDKGLPKHLELDKSHLEGLRKPQD-----PVKLGEBAHLPHL 2106
Db      1369 GLPGLPGKLGEBPLGPGFPGQKGETGPGQ--PGIPGLPMKMGDSGVGAPRGDAGPGKQ 1426
Qy      2107 RPLPSQSSSPRLQT-----ARGVKGHGVVTLAQHISVITITODYTRHHHPQOLS 2156
Db      1427 BGGEMGPPEAOPITVORGEKGEWPGAPGIRGK----- 1460
Qy      2157 APLPAPLYSFPGASCPVLDIRRPSDLYLRPPDHGAPAR-----GSPHSGKRSPEPNKT 2212
Db      1461 -----GLPGDGLPGPSG--PPGFAKAGKGDGPGQPGMGKGEKAP----- 1499
Qy      2213 SVLGG--GHDGIEPVSPPGMTPEPGHSRSAAVPLLYRDG---EDTSPSRMGSKSPGNTS 2266
Db      1500 ----GLPFPPIEGIPGPGPLPGPSGPPGPGP--SYKQGFLLVKHSQTSSEVYQCFPMVK 1554

```

```

Qy      2267 QPPAFFSKLTESNSAMVSKQ-----EINKKLTNHRNEPEYNI 2306
Db      1555 LMDGSLTYIEGNE---KSHNODLGHAGSCLSFSTMPPLPCDVNNVCYASRNDKSYVL 1611
Qy      2307 SOPGTEIFNMPAITGTGLMTYRSGAVQEHASTNMGLEAIRKALMKYDQWEEPSPLSAN 2366
Db      1612 ST--TAPIPMFVSSGIEPIYSRCVACEAPAN-----VIAVHSQTIQIPN-CPN 1658
Qy      2367 APNPPLNASSLSLPAAPITRADORSDHTLTPGCGGKAKYSGRPSRRKAKSPAPGLASGR 2426
Db      1659 GNNSLMTIGYSF--AM-----HTGAGBGGQOS-----LSSPSGCEDFRA 1696
Qy      2427 PPSVSSVSEGEDCN 2440
Db      1697 TPPIECNGARCTCH 1710

RESULT 45
T31421
C-terminal domain-binding protein ral - rat
C1:Species: Rattus norvegicus (Norway rat)
C1:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C1:Accession: T31421
R1:XuYuev, A.; Patirajan, M.; Litington, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A1:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit
A1:Reference number: Z21024; PMID:96293459; PMID:8692929
A1:Accession: T31421
A1>Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: mRNA
A1:Residues: 1-1173 <YUR>
A1:Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AA052657.1
A1:Experimental source: h1pocampus

Query Match
Best Local Similarity 21.5%; Pred. No. 4.2e-06;
Matches 311; Conservative 154; Mismatches 537; Indels 447; Gaps 69;

Qy      753 SPHTAAKDQNGKPPA-----TLGADGPPPPPPPPRTSPAPLEPTPASATGAP 806
Db      77 SPQSHASR-----PACSRHLTLTGDCGPAPPPAPSSGSSSP-SPSPSSSPBP 126
Qy      807 TPPAPPSAPPP---VVPKEKEEETAAPAVEEGEQKPPABEIAVDGKAEPPVK 863
Db      127 PPPPPPPALDAPRFDIYDPHPTDEAVSPPAEQKYDPEALGSPSSSGTSPSE 186
Qy      864 SECTEABEGPAKGDAEAMATAGALKAEKKGSGRATTAKSGAPQDSSSATCSA 923
Db      187 EEEEBEEEE-----EEGLSQSIRRISETLAGIYDMSLSQDPFG 225
Qy      924 DEVDAEGDKKRLISPPRSLLTPGDP-----RANASPOKPLDKOLKQRAAIIPIQ 977
Db      226 DDSPHRE-----PPPPQTLGAPGTPOADSTABEAGPRRRVVVVGPEAEACLEGKVS 277
Qy      978 VTKVHEPPREDAPIKPPAPAPPPQNLQPSDARQ--QPGSPR-----GKSRSAPAP 1029
Db      278 V---EVVTTAGGAPLPPLPDP---TDPELEBESEIVQPEEBEFVAVSLRAAPRQDP 329
Qy      1030 ADKEAFAAEQCLPGDPPCWTSGLPFPVPREVIKASPHADPSAFYAPGHPPLGLH 1089
Db      330 ASVATLASVA--APAAPA-----SAPRAPGDDF-----LSLH 361
Qy      1090 ----DTRPV-LPPPTITSNPPPLISSAKIPSVLERQIGALISQMSVQLHVPYSEHAKA 1143
Db      362 ADSDEGALQVLDGEPRA-----PPAADARWGGLDIRRKI-----LTQREERYQORAS 410
Qy      1144 PVGPVTMGLPLPMDCKLAPSGVQKQLSPRGAQGPESIGVPTAQ--EASVLRGTALGS 1202
Db      411 P-GP-----PPARKKR-----RERQRS--GDPAPDS--PLWEAKKGRSRERKIGS 452
Qy      1203 -VPGSITKGIPISTVPDSAIYVRSITHTGTADVLYKGTITRIIGDSSPRLDRGED 1261

```


A:Reference number: Z20787; MUID:96431717; PMID:8834799
 A:Accession: T30249
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2938 <STA>
 A:Cross-references: EMBL:X82786; NID:g1177527; PIDN:CAA58026.1; PID:g1177528
 A:Experimental source: strain CBA; testis
 C:Genetics:
 A:Gene: KI-67
 C:Keywords: cell cycle control; nucleus; tandem repeat

Query Match 2.8%; Score 373.5; DB 2; Length 2938;
 Best Local Similarity 19.4%; Pred. No. 1e-05;
 Matches 576; Conservative 351; Mismatches 1073; Indels 973; Gaps 145;

QY 43 QHNSDYASHLPGSIIPORRRPSILSEFQGNER--SQELHLRPSHSLYPLKQSE 99
 DB 182 QHTGNRIIVEPTSGGSL-----RSPGLQAV--TGNRSLPTQSLSNSNEKESPEKLYQGM 236
 QY 100 MEFIESKR-----RLLELDPPL--RSPPLATG---QPAQSEDLTKDRS-LTGKL-- 145
 DB 237 KEELDYKQSKCRKSPQPDRAAESRETQLLVSGARAKSSGSTFTVTAAASPVKGKIWT 296
 QY 146 -----EPVSPSPPTDPELEIVPRLSKEELIQNMDR-----VDREITM 185
 DB 297 EKRGMGVVQSTETAKKKTTPRHQOLKDEDSRVTGRHGVNLDGGSSAQAVHKTVTP 356
 QY 186 VEQIISKLKKQOOLEE--EAAKPE--PEKVPSP-----PRIESKR-SLVQIIT 231
 DB 357 -----GKLATRNQTPVADVGSPADTPHSSSPQASIPAKVEAPAFQNTLSLTQRLV 411
 QY 232 DENRKAEAHRLBELGQVELPLYNOPS-----DTRQYHMKIKINQARKKLILYFKR 287
 DB 412 PEKKTPKSPSKPELATPAEQTCGLFGLSSVDLSNFGDSINNSEGMPMK-----RKR 466
 QY 288 -----NHARKWKQKFCQRYDQMLEALEKVERIENPRRRAKESKREYEEKOFPEIRKQ 343
 DB 467 VSRGHLRP-----ELFDENLP-----PMTPLKRGFTPTKSKLGTSHSPAVLK- 509
 QY 344 RELQERMOS-----RVQORGSLSM----- 363
 DB 510 TIIKERPOSFGKQSPGITEPTPTNDQRRSGRTSSGNFLCETDIPKAKRSGNLPAKR 569
 QY 364 -SAASEHEVSEI-----DGLSEQENL-----EKOMQOLAVIPMLYDAD 403
 DB 570 ASISRSQHGLQMTCSKRSSGASEANLIYAKSWADVVKLGVCQQTQTKVAKHVPKQTSKR 629
 QY 404 QOR-----IKFINMGIMADPMKVYK----- 424
 DB 630 QRRPSTPKKPTSLNHQFTTGHANSPCTIVGSAQILEKVS-----PARPYMLNNLML 683
 QY 425 DRQV-----MNMWSEQEKETFEKFMQHPKRNGLI--ASFLEK-----TVACVLV 469
 DB 684 NRKVDSEDLGSLTEFMKTPFVKEKQOQMSDTSVLNSANLSEKQLQVNSGDIPRPIYT 743
 QY 470 YVLTIKENYKSLVRSYRRRGSGQQQQQ-----QQQQQQQQQQQPMRPSSEKDEKX 524
 DB 744 EILIGER-----VLSTRMAAQSDRYSASPTLKRRIKHNNTVOTPRKVNITDLEK 796
 QY 525 KEKEAKEKEKEPEVENDKEDLLEKXTDDTSGENDKEKAAVASKRKTANSQGR--RKGRI 583
 DB 797 KIPVSTETPLKTSVSKLRSRELHNTLVETNMKEATEVLAENTTAPHLRGTFRQKVD 856
 QY 584 RSMANRANSEALTPQOASALAMEINNESSRMTBEMETAKKGLLEHGRNMSAIAMVGS 643
 DB 857 QQVODNENA-----PQRCKE--SGELSEGEKTSARRSSARK-----QKPTKDLIGS 901
 QY 644 KTVSQCKNPFYFNKRR-----QNLDELQOHKLMKEMERARRK----- 683
 DB 902 QMVTQTDVADVAEELLISQGGTIOQLLESWMQNTSISBDGITEKVNIIYATKEKHSK 961
 QY 684 ---KKA-----PAASEEAAP-----PVVED-----EEMASGVSG 712

DB 962 TPQKKAQPLEGPAGLKEHFETPNPKDPEITEDRRVLCKSPQVNTENTITNTKQSTSG 1021
 QY 713 NEEBVEAEAL-----HASGN-----EVRPGCSG-----PATVNSSDT--- 748
 DB 1022 KKVDMKESSALYTKRIHPGSRHNPKILKECEDIKALKQSENEMLTSTYNGSKRTIGK 1081
 QY 749 -----ESIP--SPHTE-----AAKDTGONG-----P 767
 DB 1082 SKKKAQPLEDLTCFQELFISVPPTNI IKKITSKSPHTQPVRTASTKLSKTLGSKVDVR 1141
 QY 768 KPPATLG-----ADGPPGPPPTP-----RTSRAPRPT 797
 DB 1142 QEPSTLGRKTKSPGAPGTPPAVQENDCTAMETPKQKLESIENTLGLRQOSRPPKQIT 1201
 QY 798 PASATGATPPRAP-----PSPSAPVPVPEKEEBEETAAPAVEEGEQKRP 846
 DB 1202 GFQDSFQIPDHANGLVVVKTKQMFNSPQ--PESAITKSRSGRSAS-----ISKID 1253
 QY 847 AAEBLAVDTGKAEBPVKSECTEAEERGPAKGDAAEATAGALKAEKKGSGGATTA 906
 DB 1254 VKEELL-----ESEHNLQLEGVDTFQVSTNKVIRSSRKPAKRLDSTA 1297
 QY 907 KSSGAPQ-----DSDSATGSAD-----EYDEABGGRKRLSPBSLL--TPRGDP-- 951
 DB 1298 ---GMPNKRKMKSCSKDNTPLIEDLNGFQELFQMPGVANDSLITVGI STMARSPLQGVPR 1354
 QY 952 -RANSPQKPLDKOL-----KORAAIPIQ-----VTKVEPPRED---AA 990
 DB 1355 TQINKKSLPKTILRKMDVTEELSGMKOSLGRVHTTQOEDNAIATHEIPKETIQTAD 1414
 QY 991 PTKPAPAPAPPPQNLQPSD-----APOQSGSPKGRSPAP-----PA 1030
 DB 1415 GTRLRQGTPEKEXQPLBDSHFQELFQTSRYCSDPLIGNKQTMSLRSRQGVHTPR 1474
 QY 1031 DKEAPA-----AAQKLP--GDRPCWTSGL--LPPVPPRE-----VTKSPHAPDDSA 1074
 DB 1475 TSRLAKTSVGNIAVREKISPVSLPQCATGEVNHPIPEBDTENKGVKSESTPQTLDSA 1534
 QY 1075 -----FS-----YAPGHPRLPLGLHDTARPVL--PRPTISNP 1105
 DB 1535 SRTVSKRQOGAHEERPOFSGDLFHPQELFQTPASGKDYVYUDETIKIALQSPQHIIINP 1594
 QY 1106 P-----PLISSAKHPSVLERQIGALISQMSVOLHVPYSEHAPAVGVPTWGLPLPMD 1157
 DB 1595 ASMKRQSNMSLAKDMRPSILEKQ--TQSRGDA-----GTPAPMQ 1633
 QY 1158 -----PKLAPF-----SGVQOQLSRGQAGRPESLGVPTAOEASVLRGTALGS 1202
 DB 1634 EENGTATIMETPKOKLDFIGNSTGHRPRRTPKNAQPLEDL--DGFQE--LFGTPAGA 1688
 QY 1203 VPGSITTKGIPSTRVPSDS-----AITVGSITHTPADVLVYKGTITRIIGEDSPSLD 1256
 DB 1689 SDPVASEEAKISLASSQAEVPRTASTKRRKTLGSKVDV-----RQSPSTIG 1737
 QY 1257 RGRBDSLPGKHVIYEGKKGHVLSTYEGMSVTQCSKED--GRSSGGRPHETAPRYTYDM 1313
 DB 1738 K-RMKSIGRA--PQTPAPVQENDSTAFMETPKOKLDFNTSGSGHKKRRPQTPKIRAQ 1792
 QY 1314 MEGRVG-----RAISSASIEGLMGRALIPREHSPHNHLEQHNHIGSITQGI PRSVV 1364
 DB 1793 LEDLDGFOELFQTPAGANDSVIVESVKKSLSSQAEF--VTPPASSTKLSTGL--SKV 1848
 QY 1365 EAOEDYLAREAKLIREGTPPPPSRDLTEAYKT--QALGFLKLPRAHEGLVATVKEAG 1422
 DB 1849 DVREDSIILEK--TSPGTPAVQEBNDCTAMETPKOKLDTYGNSSGKKRRPRTPK 1904
 QY 1423 RSIHEIPREELNHTPELPLAPRLKESIT-----QGTPLKYDTGAST----- 1465
 DB 1905 --IRAQPLEDLGFOELFQTPAGASD--SVTVESAKMSLESQAKPVK--TPASTKRLSK 1959
 QY 1466 TSGKXHDV--SLIG-----SPGRTPPPNHPLDVADAPALRACYEESLKRPGTASS 1518
 DB 1960 TGISKVDVREDSITGKTKTSPPRA--PQTPAPVQEN--DSTAFMETPKOKLDFEANS 2014

atrophin-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
 C:Accession: S50832
 R:Nagafuchi, S.; Yanagisawa, H.; Ohaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.; Yam
 Nature Genet. 8, 177-181, 1994
 A:Title: Structure and expression of the gene responsible for the triplet repeat disorder
 A:Reference number: S50832; MUID:95144175; PMID:7842016
 A:Accession: S50832
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1184 <NAG>
 A:Cross-references: EMBL:D31840
 C:Genetics:
 A:Gene: GDB:DRPLA; B37
 A:Cross-references: GDB:270336; OMIM:125370
 A:Map position: 12p13.31-12p13.3112p-12p

Query Match 2.8%; Score 373.5; DB 2; Length 1184;
 Best Local Similarity 20.0%; Pred. No. 3.9e-06;
 Matches 329; Conservative 128; Mismatches 588; Indels 603; Gaps 73;

QY 520 KBEKEKEAEKEEKEPEVENDKEDILKKTDTSGEDNDEKAVASGKRTANSQGRX 579
 DB 2 KTRONKDSMSGRKKEAPGREELRSGRASPGGVSTSSSDGKAEKSGRTA-----KK 56
 QY 580 GRITSMANNEANSEAITPOQSAELASMLNSSRWTEEMETAKKGLLEHGRNSALR 639
 DB 57 ARV-----EENSTPKVKKQGRSEIISSE----- 80
 QY 640 NVGSKTVSCKNFYFNKKRQNLDELQHKLMKEKERNARRKKKAPAAASEEAPPV 699
 DB 81 -----SETNAPKTK-----TEDELPRP 99
 QY 700 VEDEMEASGVSGNEBEWEAEALHASGENVPRG-----ECSGPATVNNSSDTE 749
 DB 100 QSPSLDLSL-----DGRSLNDGSSDPRDIDQDNRESTSPISYPCGVENDSDSS 148
 QY 750 SLPSPHTEAKDTGNGPKRPATLGADPP--PGPPTPRTRSP--IEPTPASEATG 804
 DB 149 SGLS-----QGRAP--YHPPPLFPSPQPDSTPRQDEASFEPHSPVTPTG 193
 QY 805 --APTPP-----APSPSAP-----PVPVKEKEKEETAAAPVEBGE 841
 DB 194 YHAPNEPPTSMFQAPGARPPHPQLYPGGTGVLSGPMPGKGGGAASVSGP---NGG 250
 QY 842 ECKPRAAEELAVDTGKAEEPPVSECTEAEESGPANQDAEAALAEALKAKEKGGSG 901
 DB 251 KQHPPTPTTIVSSSGA----- 267
 QY 902 RATTAKSGAPQDSDSATCSADEVDEAGDKNLLSPRSLTLPTGDPANNASPQKPL 961
 DB 268 -----SGAPPTKPTTPVG-----GGN-----LPSAPPANFHTVTPMLPPP 305
 QY 962 DLKOLKORAAALP---IQVTKNHP--PRADAAPTKRAPRAPRPONLQPSDAPQOQS 1017
 DB 306 ALRPL--NNASAPPGIAGQPLRGHLPSPYAMGQMGGLRPGREKPTLAP-----SHS 358
 QY 1018 SPRGKSRSPPAP---ADKFAFAEAOGLPDDPCMTSGLPFPVPREPVIKASPH--PD 1071
 DB 359 LPPASSSAPAPMRPRPYSSSSSSAAASSSSSS-----SSSSAPPPASQALPSYPHSFP 414
 QY 1072 PSAFSYA--PCHPLPLGLHDTARVPLR-----PPLISNPP-----LISSAKHPSVL 1118
 DB 415 PTLISVSNQPKY-----TOPSLPSQAVMSGPPR--PPYGRLLANSNAHPGF 462
 QY 1119 EROLAISQGSVQLHVPYSEHAKAPVGVTVTGLPLPMDPKLAEFGVKOEOLSPRQA 1178
 DB 463 PPSGAGSTA-----HPVSTH-----HHHHQOQOQOQOQOQOQOQOQOQOQO 502
 QY 1179 GPPESLGVPTAOEAS-----VLKGTALGV--PGGSLTKGIPSTRVPSDAITTYRG 1227
 DB 503 GPPPPGAPRHPLEGSSSHHAPRYAMSPSLGSLRPYRG-----PAHLRPHSQVYSQ 555

QY 1228 SITHGPAVDVLYKGTITRIIGEDSPRLDRGEDSLPKGHV--YRKGKGVLSYEGGMY 1286
 DB 556 AAPPNGPPVS-----SSNSSSSSTSGQ---YPCSHSPSPGPGQAPVPFPVPTV 602
 QY 1287 TD-----CSKEDGRSSSGPPHETAAPKTYDMEGRVGRALSSASIEGLMGR 1334
 DB 603 TTSATLSTLVATVASSPAGYKTASPPGPPPKR-----ASPGAYKTA 647
 QY 1335 TPERHSPHHLKEQHINKSITQIGIPRSTVEAQEDYLRRBAKLKREGTTP---PPPS 1390
 DB 648 TTP-----GYKPGSPSP-----RTGPPRYSRTSP 675
 QY 1391 RDLTEAVYKQALGPKLKRA--HEGLVATVKEAGRSIHEIPRELHNTPEPLAPRLK 1449
 DB 676 GGTGTRKPGSPYVGPPLPAGPSGL-----PSLPPPPAPASGPPLSAT 719
 QY 1450 SITQGTPLKYDTGASTGSKKIDVRSLLGSPGRT--FPVHPLDVNADARALERACYE 1508
 DB 720 QIKQPAEYETPESPV-----PPARSPSPPKVVDV--PSHASGSAFNNKHL 765
 QY 1509 KSRPGTASSSGSITARGAPVIVPELGK--PROSLTYEDHGAFFGHLPGSPVTMR 1566
 DB 766 -----DRGFNSCARSDLYFVPLEGSPWPRSGPTWRRCGAGRAARAR----- 808
 QY 1567 PRLQSGSLSSSKASQDR---KLTSTPRELAKSPHSTVBEHHHPISPEHLLRGV 1623
 DB 809 ---KGAARABREKEREKERDYERSVKLAQEBRAVECB--SIGVPVH----- 853
 QY 1624 YRSHPLAFDPTSLPRGIP-----LDAAAYLLPRHLAP--NPTVPHLYP-----PY 1668
 DB 854 -----RPPPEPSAAVATVPYLGPDTPALRTLSEVARRHWSPGRNHPFVPLGAV 908
 QY 1669 LIRGY-----PDTALENQTITINDYITSCQHHNTATAMAQADMLRGISPRESS 1721
 DB 909 LL-GYNVALYSSDPAERERE-----AREDDLRDLKP----- 942
 QY 1722 LNYAGPRGIIDLSQVPHPLVLPPTGTPATMDRLAYLTPAPQFSRSSSPSLSPG 1781
 DB 943 -GFVKESEL-----BPLHVGPGGLD-----PF--PRHGGLAQGP 977
 QY 1782 PTHLTKPTTSSSERERDRDRERERKSILSTTVHAPRTWPRGTEOSSSGSSS- 1840
 DB 978 PGLHFPFPHSLGPIERER-----LALAAGRLARDMVAEYELAER 1019
 QY 1841 -----SGGGGSSSR-----PASHSHAHQSPISFRDQALQRPSTLHNTGKIT 1887
 DB 1020 QHAERVAGLGNDPLARLQMLNVTPHNHQSHISHLHLOQDAILHMASASYPL----- 1073
 QY 1888 TAVERSKPTVLASTSTSPVRPATFPRAATCPILGGLIDGYVP--TLMEPVILPREAPRV 1945
 DB 1074 -----IDPLASGHLTIP-----YPAGTLNPP--LLP----- 1099
 QY 1946 ARPERPADTHAFLAKPPASGLEPASPSKSGSPRLVPVSGHATARTPAKN---L 2002
 DB 1100 -HPHENVLHQLFAAYRDL---PASLSAPMSAAHQLMHMQSALQLALQEQQWL 1155
 QY 2003 APHNASPPRAPPASASDPHREKTQSKP 2030
 DB 1156 HAHNPLHSPVLPDAQEDVYSHLKESDKP 1183

RESULT 42
 T14355
 Protein-tyrosine-phosphatase (EC 3.1.3.48) TD14 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14355
 R:Cao, L.; Zhang, L.; Ruiz-Lozano, P.; Yang, Q.; Chien, K.R.; Graham, R.M.; Zhou, M.
 J. Biol. Chem. 273, 21077-21083, 1998
 A:Title: A novel putative protein-tyrosine phosphatase contains a BR01-like domain and s
 A:Reference number: Z18004; MUID:98361981; PMID:9649860
 A:Accession: T14355

Best Local Similarity 19.0%; Pred. No. 8.8e-06;
Matches 388; Conservative 249; Mismatches 799; Indels 607; Gaps 84;

```

Qy 15 TERYPHSLSTVPOIARTHTDVLGLVYOHHSRDYASHLSPGSIQ--PQRRRLSLSEF 72
Db 846 TMIYVPP-----PHLIAATSTGT-----HSVSSASHSTPRHSISGTPVCEPSNSKTS 893
Qy 73 QPNERSGEHLRPS-----HSYLPELGKS-----EMEFIESKRRLLELPPLLRP 120
Db 894 QPPEPKS-----RPEKVOIRHDTTISKSGPSNAINALQARSQSMSTSGDKKAPSTPVVD 948
Qy 121 --SPILA---TGOP--AGSEDLTKORSLTGKLEPVSPSPPTDPPELPELPRLSKEELIQ 174
Db 949 AGSDLVAGIMNQPLGLRLKPRIEKSSALQNIQNHPPHSNANSTSTSTSTHQMFM 1008
Qy 175 NMDRVREITWVEQOISLKKKKQOOLEEAAKPEPEKVPSPPIESKRSLLVQIYDEN 234
Db 1009 K-----DK-----EKRRKKKEKEKEERREARR-----EMRKETKE---BRN 1043
Qy 235 RKKAABARITIEGLQPOVELPYNOPSDTROYNENIKINQAMRKLLIYFRRNH----- 289
Db 1044 KRKEMERAKLEDERQERKREKREKDERKKEKVR--KKAKEKL-----KKKKRKGDS 1098
Qy 290 -----ARKWKOKFCORYDOLMEALEKKEYERJENNPBRRAKESKYREYE 334
Db 1099 SDESDSDNDELDDVYRKSTKEMTOEKKDQHLALLSGGIIENLKSRRSDKRAHDSFE 1158
Qy 335 KQPEIRKQRELOEEMQSGVQCG--SGLSMAARSEHEVSEITIDLSQENLEKQMROLA 393
Db 1159 KMOQSGQSRVLISSDEGCKDQKNSNGEESDEKADL-----PPP 1203
Qy 394 VIPPLVYADQORIKFIMNGLMADPMYVYKQROMNMSSEBKEKTPREKFMQHPKNGL 453
Db 1204 PAPPLSSSAQOR-----LVYKEREKGLTSSDDHDNDAGEIHOQR--- 1247
Qy 454 IASFLERKVAECVLYYLTKNENYKSLVRSYRRKSGQOQOQOQOQOQOQOQOQOQOQO 513
Db 1248 -----LTEDREN-----RKQKSLTAYSSDQGERKANVPKMR 1280
Qy 514 RSSQEE-----KQEKERK--EAEKEERKEVENDKEDILKEKTDPTDGEDNDE 560
Db 1281 RDSSEDAAKHPGMAKDDQKQKRLKRRRSEDESCKNAKRDPRDIPEHVDSEBEETE 1340
Qy 561 KEAVASKRK-----TRANSQGRKGRITRSMANSESEAITQ--QSAELASHEL 609
Db 1341 D--GSRRRQSTSTISNTVAKERKESGKTPRLIPEPTGPLLSPKSLSPKLSKPT 1397
Qy 610 NESSWTBEEMETAKKGLLEHGRNMSAIARWGSKTVGCKNFYNYKKRQNLDELQOH 669
Db 1398 STSS-----TKRSSISDH-----EULISPRQRN 1420
Qy 670 KLKQEKERNARRKKKAPAAASEEAPPV-----VEDE---EMASGVSG 712
Db 1421 R-TTSSSTATTATSSGHEALSIPEKPLSPVTAKSSVSSIDPSIRDFSMNSAASPST 1479
Qy 713 NEEENVEAEALHAGNEVPR-----GECSPATVNNSSD--TESIPSPHTEAKD 761
Db 1480 TGRPVNLTAKAMKAKNSTPPKKKNSSSGQHDSSSSSSSDSSDSTSDSDSDDEVKQ 1539
Qy 762 TGQNGPKRPATLGAQPPPPPP-----TPPRTSPAPIEPP--ASEATGATPPAPPS 814
Db 1540 T-----EYVTSIPVVASDNGSPENVVETPSIVGOTPREPEFTTISEGSSBEPFAVEBC 1594
Qy 815 PSAPPVVPKEEKEETAAPRVEEGEQKPPAABELAVDTGKAEVPVKSCTEEBAEGP 874
Db 1595 PEA--SVPEQMETSON--VEPVSEHDSHEHGDSEVAVES--QOQPLEHQ--EKEELE 1646
Qy 875 AKGQDAEA-----EATFEGALKAEKKGSGGRATITAKSG--APQSDSSAICS 922
Db 1647 NKILVAAEHHEEQVQGDSDVSSSIPAPSDPEPVTVOAQEKSAHTLLISDQTDVAQSI 1706
Qy 923 ADEVDAEGCDKNRLLSPPSLLTPTG-----DPRANASQKPL---DLKQLOKQRA 970

```

```

Db 1707 FDEBEADE-----PQYDPFGISTNEKEVSGKDPH-NIKTEPLNNGHTLLFSBSSS 1758
Qy 971 AAIPIQVTKYHEPRRDAAPTRKAPAPPPQVLQESDAPQOQSSPRKSKSPAPPA 1030
Db 1759 AAASEKQSTSEDMEDSESLVNEKEV--PHEQVIAQEVHVPSHP-----SPM-- 1805
Qy 1031 DKEAFAAEQAQLPDPDPCMTSGLPFPVPPREVIVASPHADPSAF-----SYAPPG 1081
Db 1806 -EEVKLETSPVPEEPPIKHEESPEQTPTRLINNNSQDTPGAVNNHLENHNAVQPI 1864
Qy 1082 HPLPLGLHTAPVLPVPPPT-----SNPPLISSAKHPVLERQIGALISQ----- 1128
Db 1865 QLOPASQHOVAP--SPRPVAVPDSQNGPVLVSOQSPSPMSQOQDMAQVLLISSKDIN 1923
Qy 1129 -MSVQLHVPYSEHAKAPVGPVTMGLPLPMPDKLAPPSGVYQEDL---SPFGQAGPRESL 1184
Db 1924 DLAAKLHNPEALAOATRQDC-----SGIFQHLLLHAQNGQNMTPBML 1967
Qy 1185 GVPTAQASVYLRGTALDSVPQGSITKGPSTRVPSDAITVRSGITHTGPADVLYKGTIT 1244
Db 1968 QLKAAFPQOQOENALNQMAQKQOITINDRIKEQERVAK-----MYEEN-E 2014
Qy 1245 RIIGEDSPSRLDGRBDS----- 1263
Db 2015 RKVEED--REKQKKEERQRLAATAATATQKAAEALQKQEVPRHGFQVLSMT 2071
Qy 1264 PKGHVITBKKGHVLSIEGMSV-----TQCSKEDRSSGPPHEHTAAPRT 1310
Db 2072 PEARSLYEOPFG--LSSYINRDSIGATNGVLLPTQSIORPSTASTSNPKAPLOPAS 2130
Qy 1311 Y-----DWMEGRVGR--AISASIEGLMGRAPPEBSHPHLKEQHIRGSIQOG 1358
Db 2131 VNQNTIPDAIEIEIRVQRMFKPLKMSAEBAATMAVASSPNP-----PATSYVD 2181
Qy 1359 IPRSYVAQEDYLREAKLKRGEPTPEPPSRDLTEAYKTQ-----ALQPLKL- 1407
Db 2182 LAAMLQQLQAAQAAQAAQVAVVTASTPNLSLETLLSTASLANLATGALNPLSMILA 2241
Qy 1408 -----KPAHEG-----LVATVGE-----AGRSIHETPR 1430
Db 2242 LTSSLNQSSPYQGIARVLLTMNNGOMLATHTQSELTATMNOBTMALLAARNGLPFAM 2301
Qy 1431 EELHHTPELP-----LAPRLKEGSIQGTPLKYDTGASTGSKHDVRSIGSPRTFP 1485
Db 2302 PQQNOQPMFQAGGFAIPVTLPHMSIKRNADQLSVGVSBRKKSCLPHAMITGGQOQPP 2361
Qy 1486 FVHPLDVMAADARALERACYEBSLKSRRGTASSSGSIAARGAP----- 1527
Db 2362 PQQPMQAVAPA-----PPRSPPPPKSMFENLPPEWKEKNEMFRKEILRL 2407
Qy 1528 --VIVPELG-----KP--ROSPLYTD-----HGAPPAGHLPRGSPVTAREP 1565
Db 2408 DIIILIEEGADEBDQKDLQIPTSSEDTDSKADSGAGSAFRRLTSSS--TM--- 2462
Qy 1566 TPRLQEGSLSSKASQORKLSTPREIAKSPHSTVPEHHPRI--SPYEHL-----RG 1617
Db 2463 -----GNNGSPBASGTTSTSTSSISSGDS--PRLEGPLNSBFMDLTVEVAQKRE 2514
Qy 1618 VSGVDLYKSHIPLAFDPTSPRGIPLDAAAAYVLPRLHAPNTPVHLVPLYLRYGPDTA 1677
Db 2515 QSNTDALSAKI---VDEGSEFQHFPMWTGRLAKSTEA-----MINHLING----- 2559
Qy 1678 ALENRQTIINDYITSQOMHNTATAMORADMLRGLSPRESSLALNTAAGRGIIIDLSQV 1737
Db 2560 ---SETFLNDVLRQVTEEN-----PRDSVXIL---ORLRLDNGQV 2595
Qy 1738 PHL 1740
Db 2596 EHT 2598

```

RESULT 41
SS0832

A:Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5
monomer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1
F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status F
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F:42-1462/Region: interrupted helical
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F:125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F:1527-1533,1638-1644/Disulfide bonds: #status predicted
F:1522-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 2.9%; Score 377; DB 1; Length 1691;

Best Local Similarity 20.4%; Pred. No. 4.3e-06;

Matches 375; Conservative 136; Mismatches 647; Indels 678; Gaps 87;

QY 764 QNGKPPATLGADGPP-----PGPTPPRTSRATIEPTPASEATGATPPAP-- 813
DB 143 QGPPGPPGIPGMKGGFSSIISSLPFGKNGPYGPPGPIQGLPGTGIIPGPPGPPGL 202
QY 814 -SPSAPRVV-----PKEKEETAAAPVEEG--EOKPPAAB-----L 851
DB 203 MGPPEPGLPGPKGMNFGQPKGKEQGLQGPFGGQISEKRIIDVFQGDGL 262
QY 852 AVDTGKAEBPYKSECTEBAERGPAGKDAEAATAGALAKKGGSGRATTAKSSGA 911
DB 263 PGDRGPPGP-----GIRGPPPGGEGKEGQEGEPKRGKRGDNGQGL---PGL 314
QY 912 PODSSATTCADENVDEEGDKN-----RLISPP----- 942
DB 315 PGDPYGPBGPBGD--EKQKQKDTGPPGPPGLVTPRTGTTGEGKNIQLPGLPGEKER 373
QY 943 -----SLNPTGDP-----RANASPOKPLDK-----QLKORAAI 973
DB 374 GPPGIGGPPGLGPPGGAAMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 433
QY 974 P-PIQVTVHBBPPRD--AAPTKPAPAPPPONLOPE-----SDAP 1012
DB 434 GPPPPAPAPPIIPSPDEICEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 493
QY 1013 QQPG-----SSPRGKRSAPPADK--EAPAAEAQKLPGRPPCTSGLP--P--VPPREVI 1063
DB 494 GQPGPLGPPGPPGSLGFGQKGEKQAGATGPKGLPIG--GAPAPGFPGSGKBPGL 551
QY 1064 K-----ASPHAP-----DPSAFSY--APGHPPL 1086
DB 552 TTPPKGKGKGLSGSGAGPLPGLPCTPGODGLPGLPGRKGBGGITTFGKGRPPANPGL 611
QY 1087 GLHDAFARVLPRPTISNPPLISSAKHPSVLERDIGAISQGMVOLHVPSYEHAKAVG 1146
DB 612 GLPKNIGBM--GPGFGPPG-----VGEKGIQVVA--GNPQPGIP-----GPKG 653
QY 1147 -----PYMGLP-----LPMDPKLAFFSGVQKQQLSPROAGPPESLGPVT 1188
DB 654 DPGQTTTPGKRGKLGNGRGDVGLPDP-----GLPQPGPLGPIG 696
QY 1189 AOEASVLGTLGSPGGSITKGIPTSRVPSDSATITVRSITHTPADVLVYGTITRIIG 1248
DB 697 SKGEGIGIGLPGPPG---PKGFGIIPGPPAP-----GTFKIGLEG----- 737
QY 1249 EDSPSRLDRGRDLSLPKHVITVEGKGHVLSTYEGMSVTQSKEDGRSSGPPHETAAPK 1308
DB 738 -----PPGPPGPPGPPG-----EPGFALPGRPPGPPGLP- 765

QY 1309 RTYMMGKRGVGRATISSASISGLMGRAIPRHHSPHNLKEQHNIGSITGIPRSYVEAQE 1368
DB 766 -----GFKG--ALGPPGDKGPPGPPPP----- 786
QY 1369 DYLRREAKLTKRECTPPPPPPSRDLYAUYTQALGPKL-----KPAHEGLVATVKEA 1421
DB 787 -----GRITLDGLPGPKDVGPNNGQGRPMGPPGLPGLVQVGGPPGPIGPIQGP 836
QY 1422 GRSIHEIPREE-----LRHPEPLAARPL--KEGSTIQGTPLKXTDTGAST 1465
DB 837 G--LHGIPGEGKDPGPPGLDVPPGPPGSGKIGAGAPGIPGPPS--PGLPGK--AGASG 890
QY 1466 TSGKHVRSILSGPPGTFPPVPHLDVADABALERAQYESTLSRPGTASSGGSIARG 1525
DB 891 PPGTKGEM--GMMGPPG-----PPGLGIPGNS--GVPLKGDGLQGLPGLPPTGPKSKG 944
QY 1526 APVIVPEIGKPPQSLTYEDHGAFFAGHLPRGSPVTVREPTPLQEGSLSSSKASODRKL 1585
DB 945 EP-----GLP-----GPPGPM-----DNLGSKKEGK-- 967
QY 1586 TSTREIAPKSPHSTVPEHHHPHISPYHLLRGVSGVDLYNSHTLADPTISIPGIPIDA 1645
DB 968 -----EPGLPGIPGSGPKGYQGLPDP-----GQF--GL 995
QY 1646 AAAYTLRHLAP--NPTYPH-----LYPPYL-----IRGVDTALENQTI 1685
DB 996 SGQGLPGRPPGKPNPGLPGQPLIGPGLKGTIGDMGPPGQVGEPPGSGVPQGP-- 1033
QY 1686 INDYTSQOMHNATMAQPADMLRGLSPRESSLALNYAAGPPIIDLSQVPHLYV-- 1743
DB 1054 -----SGPGLPGQKD--KG--DPGISSIGLPLGPRGKGBGLPVPQNPGLK 1098
QY 1744 -----VPPTPGP--ATAMDRLAYLPTAPQPFSSRHSSPLSPGCPHTLTPTTSSSR 1736
DB 1099 SVQDPGLPGLPFTGAGAQGPGLPGLPPTGPPGPKGISGP--PQNPGLPGR 1148
QY 1797 ERDRDRERDRERKESKILSTTVENAPI-----WRPTGEGSGSSGSSG--GGGSS 1848
DB 1149 -----GPVGGGHPQGPGRGPKRGQDGTGPPAQK 1181
QY 1849 SRPASHSHAHQHSISP--RTQDALQORPSVLANHTGMKGIITAVEPSKPYLRSTSTSP 1906
DB 1182 GEPQPGGNGPPGPPGLGSLGQKDGGLPGLPGRPLG-----PKBERGHNPPGVQGP 1236
QY 1907 VRPAATPRATNCPGLGTLGVYPTLMEVULLPKEARVVARPERPRADTGAFLAKPPAR 1966
DB 1237 PGPFGSPRLALEGPKGN-----PGQGP-----PGRP--GPTGQGLPGRPG 1277
QY 1967 SGLPPASPSKSGSEPRPLVPPVSGHATARTAPAKNLAPHNASPPRAPASADPHREXT 2026
DB 1278 PGL--PKNMGIKGKGNPQGLPGLPGL-----KDDQDPGLQGNPGR-- 1320
QY 2027 QSKPFSIOELRLSLGYNGSSYSR--EGVEBVSFVSSPSLTHDKLPKHLBELDLSHLEG 2084
DB 1321 -----GLNMKGMDPLGVPVFPKMGKS-----GVP-----GSAGPEG 1354
QY 2085 E---LRKQRPVYLGEBAHLPHLRPLPESQSSSPLLOTAPGVYKGNQVVTTLAQHISE 2141
DB 1355 EPGILGP--PQPLPGLG-----PSGOSIILIKGDAAPPPIPOQPLK-- 1394
QY 2142 VITQDYTHNHQOQSAPRLAPL--YSFPGASCPVLDLRRPSDLYLPPRDGADARGSPH 2199
DB 1395 -----LPGQGGQGLPGLPGRPD-----PGRKGLPFDAGARKGDPG 1433
QY 2200 SEGKRSBPENKTSVLG--GEGDGLIEVPSPPGMPTEPGHSRAVYPLLYRDEQTEPSHWG 2258
DB 1434 LFG-----QPTGRGIDGPPGPDGLQGPFGPGTSTVAHG-----LITRHSGTLD--A 1479
QY 2259 SKSPKNTSQPAPAFSKLSTNSANVYKSKQ-----EINKKL 2294
DB 1480 POCQGTQLQVVEGSSL-----YVQGNKRAHGODLGTAGSLRRFSTMPMFCNINNV 1533

QY 2129 HQ-----RVTLAQHISEVITODYTRHHHPQULASAPLPAPLSPFGACSPVLDLRPP 2180
 Db 2612 EKGDVGFMPGRGLKGEKRGVKGACGIDGKGGKGEGRPP-----GRPLAHHKGMGBPG 2665
 QY 2181 SDLVLPDPDHGAPAR-----GSPHSEGGKSPPEPKTSVVG-----GEGG- 2221
 Db 2666 ----VPOGSGAPGKEGLIGPKGDRGFPQGPQPKDDQGEKRGRTIGIGCFPGSGNDGS 2720
 QY 2222 IEPVP-----PEGMT-----EPGHSASVAPLLYRDGEQTEPSRMGSKSPGNTS 2266
 Db 2721 AGPPPPGPGSVGRGREGIGQCGKRGPPGERVVGAPGVGAGGERGEGQGRCPAGPRGK 2780
 QY 2267 QPPAPPSKLTESNSAMVSKSKKEINKKANTH-----NRHEPPYNISQPTETLFN 2315
 Db 2781 GRNA-----LTEDDI-----RGFVRQEMSOHCACQGFAGSGRPLPSYADTASQULHA 2830
 QY 2316 MPAI-----TGTMTRRSQVQVQEHASTNMGLAIIKKALMGKVDQWEEPP 2362
 Db 2831 VAVLVASHAEEERPRPEDEDESEYSEVSEYQDP-----EA-----PWDEDDP 2875
 QY 2363 ISANFNPLNAGASLPAPAMP--ITADGRSD--HTLTSPGGGKAKVSGRPSRRAKSP 2417
 Db 2876 GCL-----PLDEGSCRAYTLRWYHRAVTGSTTEACHPFVYVCGCGGANNRFGTRACERRCP 2930

RESULT 38
 S22917
 N:Alternate names: procollagen alpha 5(IV) chain
 N:Contents: collagen alpha 5(IV) chain precursor, leukocyte splice form
 C:Species: Homo sapiens (man)
 C:date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text change 21-Jul-2000
 C:Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
 R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 267, 12475-12481, 1992
 A:title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi
 n Alport syndrome patient.
 A:Reference number: S22917; MUID:92316923; PMID:1352287
 A:Accession: S22917
 A:Molecule type: mRNA
 A:Residues: 1-967 <ZHO>
 A:Cross-references: GB:M90464; NID:G180826; PIDN:AAA52046.1; PID:G553234
 R:Zhou, J.; Leinonen, A.; Trygvason, K.
 J. Biol. Chem. 269, 6608-6614, 1994
 A:title: Structure of the human type IV collagen COL4A5 gene.
 A:Reference number: A54365; MUID:94165049; PMID:8120014
 A:Accession: A54365
 A:Molecule type: DNA
 A:Residues: 1-922 <ZH2>
 A:Cross-references: GB:U04470; NID:G463378; GB:U04520; NID:G463428; PIDN:AA027816.1; PID
 R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Trygvas
 Science 261, 1167-1169, 1993
 A:title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sm
 A:Reference number: A57079; MUID:93361972; PMID:8356449
 A:Accession: A57079
 A:Molecule type: DNA
 A:Residues: 1-27 <ZH4>
 A:Cross-references: GB:Z37153; NID:G587203; PIDN:CA085512.1; PID:G587204
 R:Phlajaniemi, T.; Pohjolahti, E.R.; Myers, J.C.
 J. Biol. Chem. 265, 13758-13766, 1990
 A:title: Complete primary structure of the triple-helical region and the carboxyl-termin
 A:Reference number: A37122; MUID:90337990; PMID:2380186
 A:Accession: A37122
 A:Molecule type: mRNA
 A:Residues: 84-439, 'GS', 442-624, 'IALQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
 A:Cross-references: GB:U05558; EMBL:M58526; NID:G1314209
 A:Note: submitted to the EMBL Data Library, February 1991
 A:Note: the authors translated the codon GCC for residue 115 as Val
 R:Renieri, A.; Serl, W.; Myers, J.C.; Phlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
 Hum. Mol. Genet. 1, 127-129, 1992
 A:title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
 A:Reference number: I54317; MUID:93244772; PMID:1363780
 A:Accession: I54317

A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 313-324, 'E', 326-330 <REN>
 A:Cross-references: GB:S5934; NID:G299946; PIDN:AAD13909.1; PID:G4261609
 R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeykva, M.; Shows, T.B.; Trygvason, K.
 Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
 A:title: Identification of a distinct type IV collagen alpha chain with restricted kidne
 A:Reference number: A34850; MUID:90160375; PMID:1689491
 A:Accession: A34850
 A:Molecule type: mRNA
 A:Residues: 914-1264, 1271-1691 <HOS>
 A:Cross-references: EMBL:M31115; NID:G180824; PIDN:AA52045.1; PID:G180825
 R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Trygvason, K.
 Genomics 9, 1-9, 1991
 A:title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
 A:Reference number: A37969; MUID:91169491; PMID:2004755
 A:Accession: A37969
 A:Molecule type: DNA
 A:Residues: 924-1264, 1271-1691 <ZH3>
 A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
 8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:G177922; PIDN:AAA51558.1; PID
 R:Guo, C.; Van Damme, B.; Van Damme-Lombaert, R.; Van den Berghe, H.; Cassiman, J.J.; M
 Kidney Int. 44, 1316-1321, 1993
 A:title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
 A:Reference number: I56971; MUID:94133540; PMID:8301933
 A:Accession: I56971
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1258-1276 <GNO1>
 A:Cross-references: GB:S69168; NID:G545095; PIDN:AA060612.1; PID:G545096
 A:Note: kidney splice form
 A:Accession: I76598
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1284-1291, 'TVGTYACIV' <GNO2>
 A:Cross-references: GB:S69169; NID:G545097; PIDN:AA060613.1; PID:G545098
 A:Note: frameshift mutation in patient with Alport syndrome
 R:Myers, J.C.; Jones, T.A.; Pohjolahti, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; So
 Am. J. Hum. Genet. 46, 1024-1033, 1990
 A:title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the regi
 A:Reference number: A35335; MUID:90252791; PMID:233699
 A:Accession: A35335
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1448-1477 <MYE>
 A:Cross-references: GB:S75903; NID:G913882; PIDN:AA033374.1; PID:G913883
 R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsura, T.; Koitabashi, Y.; Takada, T.; Yos
 Kidney Int. 46, 1307-1314, 1994
 A:title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
 A:Reference number: I56975; MUID:95156893; PMID:7853788
 A:Accession: I56975
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1595-1602 <NAK>
 A:Cross-references: GB:S75903; NID:G913882; PIDN:AA033374.1; PID:G913883
 A:Note: premature termination mutation from a patient with Alport syndrome; one other mu
 R:Leimink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Trygvason, K.;
 Genomics 17, 485-489, 1993
 A:title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
 A:Reference number: I54188; MUID:94010948; PMID:8406498
 A:Accession: I54188
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1604-1607, 'VHDAYKC' <LEW>
 A:Cross-references: GB:S65767; NID:G9428563; PIDN:AA013967.1; PID:G9428567
 A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations ar
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL4A5; ATS
 A:Cross-references: GDB:120596; OMIM:303630
 A:Map position: Xq22-Xq22
 A:Intons: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/
 /3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1165/1; 1

A>Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis A>Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring F
C:Keywords: collod coll; extracellular matrix; glycoprotein; hydroxyllysine; hydroxyprol
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domain: von Willebrand factor type A repeat homology <WMA1>
F:231-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-553/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <WMA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-2783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:2876-2929/Domain: animal knitz-type proteinase inhibitor homology <BPI>
F:337-786,1109/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:2167-2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2625,2631/Binding site: carboxydrate (Lys) (covalent) #status experimental
F:2634,2802,2804/Distulfide bonds: interchain #status predicted

Query Match 2.9%; Score 383; DB 2; Length 2944;

Best Local Similarity 22.1%; Pred. No. 4,8e-06;

Matches 437; Conservative 143; Mismatches 761; Indels 638; Gaps 110;

Qy 726 ASGNEVPRGECGSPAT--VNSSDTE:IPSPHTEAKDTGONGPK-----PPATLGADGP 778
Db 1303 ADGRGSGRAGNPGTTPGAPGLKSGPLGPRGDPG-ERGRPGRGEFGAPQVIGGEGP 1361
Qy 779 -----PPGPTP-----PR-----RSTRAPETPASEATG 804
Db 1362 GLPERKDDPGSGPRGPRGLDPPGRKPPGLPGTAMKDKDKREKRPFGGEGGIAPG 1421
Qy 805 APTPPAPSPSAPRPVVP--KEEKEETAAAP-----PVEBGE--OKPPAAELAVD 854
Db 1422 EPLGLPGSPGPPGPPGPKKGSBDGAPGLPGQPSFGEGRGRPGALGPKGD 1481
Qy 855 TGKAEPPVSECTEBAEGR---AKGDAAEATAGALKAEKKGSGGATTAKS 909
Db 1482 RG-FPGPL-GEAGEGERGPPGPASRGPLGVAGRPAGK---FGPPGPTGRGEGEP 1536
Qy 910 GAPDSDSSATSGADEVDEAGGDKRLISPR-----PGLPTPDGPRANAP- 957
Db 1537 GRP--GDPAVWPVAVAGKGEKGDVGP--AGPRGATGVGGERPPGLVLP--GDPGRKDPG 1592
Qy 958 -OKPLDLQOLKRAAIPPIQVTKVHEPPREDAATKAPAPAPPPQNLQPSDAPQDG 1016
Db 1593 DRGPIGL-----TGAAGPPGD---SGPPEKGDGRGEPGP 1626
Qy 1017 SSPRGSKSPAPPADKEAFAAEAKQLPGDPCCMTSGLPFPVPREVITKASPHAPPSAFS 1076
Db 1627 VGPFRGDEVEGKEGEP-----PGP-----GLPGKAGERG-LGAPAVTGPVGEK 1672
Qy 1077 -----YAPGHPLPLGHTTAPVLEPRPTISNPPLISSAKHPSVLEQIGAI 1125
Db 1673 GDQGDPRGEDRGNGSGSGPKG--DRGEGPPGPP-----GRLVDGPGAR 1716
Qy 1126 SGQMSVQLHVHPYSEHAKAPVGPV--TMGLPLPMDPKKALFSGVKQEQSLP--RGOAGPPE 1182
Db 1717 EKG-----EPGDRQGBRGRKGPGLGAPGEGKIGBGRPPGPPQDPPGVRGAGEKG 1770

Qy 1183 SLGVPTAQEASVL---RGTLGSPVGSITK-----GIPSTR-----VPDSAIITYRGS 1228
Db 1771 DRGPGGLDNGSLGCKRGKALGPPSGNCAAGKAGDPGDDGLPGLRGEGGLPSPSG----- 1824
Qy 1229 ITHGTADVLYKGTITRIIGEDSPSRDLRGREDELPRGHVLYEBSKGHVLSYEGMSVTO 1288
Db 1825 -PPLPDKP-----GEDGKPLGNKNNGEPPDGE---DGRKG-----EKGDG--G 1863
Qy 1289 CSKEDGSSSGPPPHETAPKRTTYMMEGRGRAISSASIEMLKRALPPENHSHNLKEQ 1348
Db 1864 ASGREGR--DGPKEERGAP-----GILDPGPPGLPGP----- 1894
Qy 1349 HHITGSTTGGIPRIVYVAQEDVLEAREAKLKRGETP-----PPPPSRDL---TEAY 1397
Db 1895 ---VGPQGGPPVPGGTGTGPKDGEFGSKQEQGLPGEGLRGSGVPPVNDLLLETAIGI 1951
Qy 1398 KTOALGPLKLPRAHEGLVATVKEAGRSIHETPRELHTPELPLAPR--PLKEGSIOTGT 1455
Db 1952 KASAL-----REIVTMDSESSGFLLVP--ERRRGPRGDSGEGGPPKEGPI--GF 1998
Qy 1456 P-----LKYDTG-----ASTTSGKHNVRS-LIGSPRTFPVHPLVDVADALALERA 1502
Db 1999 PGERGLKGRDPPGPPGGLALGERGPPGSGLAGSPGKPIGLP-----GRAGVGEA 2054
Qy 1503 CYEBSLKRPPCTAASSG-----GSITAGAPVIVELKPKRPSPLTEDHGAAPFAGHLPRG 1557
Db 2055 -----GRPERGERKEGKRGQGDGPPGLP--GTP-----GPPG---PPG 2091
Qy 1558 SPVTMRPEPTPL--OEG--SLSSKASQDRKLTSTPREIAKPSHSTVENHP-----H 1606
Db 2092 PKVSDERGPPLSGEQRPGLKAKGKBPVSGSDGQPKDPRVPRGIGDRGSRGQDGN 2151
Qy 1607 PISPYEHLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAAYVLPRLHAPNPT-YPHLY 1665
Db 2152 PGLRGE--RGMAG-----PECKPGLQPPRGPFGVG-----HGDDGPGCAPGIDA 2194
Qy 1666 PPLYLRGVPDTALLENKOTIINDYITSGQMHNTATMAQADMLRGIPRESSIALNYA 1725
Db 2195 GP--AGQGPSGLKGE--GEGPPRGGLTGTGAGLGLP 2231
Qy 1726 AGPRGIIIDSGVPHLPVLPPT--PGTP--ATAMRLAYLPTAPOPFSSRHSSPLSP 1779
Db 2232 PPGSGLVPPQGSFGLPGGVGTGKRGAPRGAGKAGKODRSPPVP-----GSPGLP 2283
Qy 1780 GAPTHTLTKPTTSSERDRDRDREREKSI--LTSTTVENAP-----IWRPG 1830
Db 2284 GPVGRKBPFTGAP-----GQAVVGLPRAKGEKAGLADLVGEBG 2327
Qy 1831 TEQSSGSSGSSG--GGGSSSRPASHSNAHQSPISPTQDALQORPSVLHNTGMKGIIT 1888
Db 2328 AKGDRGLRPRGKEAGEARAEPPDPEGQAGAP--GPR-----GPKG-- 2368
Qy 1889 AVESKPTVLRSTSTSSPVRPAATPRATTCPLGCTLDGVVPTIMEPVLLPKEARPVAR 1948
Db 2369 ----DDPGV-----GVDSGPPRPPGVKKDLG--LPG-----LGAREVVGFP 2406
Qy 1949 ER--FRAD-----TGHAFIAKPARSGLBPASPSKSGSEBRPLVPPVSGHATTARTPAK 2000
Db 2407 GQTGRGEMGQPGSGERGLAGPPRGEGT-----PGLGPF-----GPPG 2446
Qy 2001 NLAAPHAS-----PDP--PAPPASADPHREKTOSKPPSIOELRLSIGYHSSISPP 2050
Db 2447 SVGPFGASGLKDKDPPGVGLPGRGERGEFGIINGEDRP-----GQEGE---P 2491
Qy 2051 EGV--EPVS-----PVSSPSLTHDK-----GLPKHLELDKSHLEGELR 2087
Db 2492 RGLTGPPSGRBERGKGDVGSAGLKGDKGSAVILGPPRPGAKADMGBERGRGIDGKG 2551
Qy 2088 PK---QPPVYKAGE-----AAHLPHLRPL--PESQPS-----SPLLQTAQVYGK 2128
Db 2552 PRGDNQDPPGDKSGKEPDKXSAGLPGLRGLLGPQGGAGAGIIPCDPSPGKDVGPGRIG 2611

```

QY 1794 SERE-----RDRD-----RERDREREKSLTSTTT-----1820
DB 2580 QDSQSNVYQNLPPVQDNLMPLPDGPKPQEDSGFRKRRPRRSARASMMFGLPILYGVRSY 2639
QY 1821 -VEHAPIMRPGTBOSSGSSGSGG-----GGSSSPASH 1854
DB 2640 GREDIPFYSSSTGKKRGRSAEQVDGADDLSTDEDDLYVYVFTRTVYSSGGEELTASH 2699
QY 1855 S-----1855
DB 2700 NIFREBQCDLPKISQDGDVDTESDTSVTATTKRSSQIPKRNKENGENTENLKIDRPED 2759
QY 1856 -----HANCHSPISPTODALQORPSVLTHTKMKGIITAVEPSK 1894
DB 2760 AGEKEHVTSSVGHKNRPMKDNCHSVSRKTKQGQDSLEAQSLSSSS-----RRHHTT 2813
QY 1895 PT-----VLKSTTSSPV--RPATTPPA 1916
DB 2814 PSDKNLMDTYNTLLKSDSDNNNSDDCGNILPSDIDMFLKNTPSMQALGESPESSSSSL 2873
QY 1917 THCPILGSLD-----GVY-----PTLMEPV-----LLPEAPRVARP 1948
DB 2874 LNLGBGLSDNSREKDMGLFVPSQQLPT-TEVDSSVSSSISABEQFLPLELP-----2927
QY 1949 ERPRADTGHAFIAKPPARSGLEPPASSPSK-----GSEPPPLVP 1986
DB 2928 -----SDLSVLTTRSPVPSQNPRLAVISDSGEKRVITTEKSVASSSDPALLSP 2978
QY 1987 PVSCHATTARTPAKMLAPHH-----ASPPRAPPASADPHREKTSKSPSSIOE 2035
DB 2979 GVD-----PTREGHMPDHFTQGHMDADHISPP--PCSGVBOQHGN-----QD 3020
QY 2036 LELASLGYNHGSYSPEGVE-PVSPVSPSLTDHDKLPKLELSDSHLEGELRPQPGV 2094
DB 3021 LTRNS-----STPGQLQVPSV--TVDIQOKVVPNSTD-----SPGSS 3056
QY 2095 KLGEBAHL--PHLRPLPES-----OPSSSPL--LOTAP-GVKNORVVTLAGHISEVITQ 2145
DB 3057 QISNNAVOTTPRHLKATEKILVANQNMQPLVLOTLPNGVQOKIQLTSSSVSTSSVMT 3116
QY 2146 DYTTHHPQ-----QLSAPRLAPLYSPGASCPVL-----DL 2176
DB 3117 NTSVSGLEPMGGGLTTLTGTLNPSLPTQSILPSPASKGLPMSHHOHLSPRATQSSFPPI 3176
QY 2177 RRPSPDLYL--PPDHGAPARGSPH-----SKGGRSP-----EPNITSVLG 2216
DB 3177 SNPFGGLIGVOPPPDPQLLVSESSQRTDLSSTVATPSSGLKRRPISRLQTRKKNKLAPS 3236
QY 2217 GGEDIEIBVSPPEGMT-----EPGH-----SRSAVVPPLYRDEQOT---E 2253
DB 3237 STPSINAVSDVSNMNTLINFPSQLPNHPSLLDGLSLMTSSHRTVPIIIRKSSIMTWE 3296
QY 2254 PSRMOSKSPGNTSQPPAFPSKLTESNSAMVK-----SKKOINKLNTHNNEPEYNIS 2307
DB 3297 PAPLLPQSVGGTAATAAGTSTISQDTSHLTSGSVSGLASSSSVLWVSMQTTTTPTSSAS 3356
QY 2308 QPGTEIFMMPAITGTGLMTVYSGAVQEHASTMGLEAIIIRKALMKGYOMESSPPLSANA 2367
DB 3357 VPGHVTTLNPKLLGPPDIGSISNLIKASQOSLGIQ-----DQVALP-SSGM 3404
QY 2368 FNPILNASALPAAAMPITAADG-----RSDHTLTSFGGSGK-----AKVSG 2407
DB 3405 FPQGLTSGT-PSITAITAASSTCVLPSTQGTGITAAASGGEADHEYQLOHVNQULASTKG 3463
QY 2408 RPSGKAASPAAGLASGDRPESVSHBGCNRRPTPLTNRYMEDRPSSAGSTPPYPPL 2467
DB 3464 IHSSQRDSDASG-----PQVSN-----FTQTV--DAPNSMG--LEQN-- 3497
QY 2468 IMRLQAGVWASPPRGLPAGSGPLAGPHHAMEEKP 2504
DB 3498 -KAUSSAVQASFTSPG-GSPSSPSGQGSASBPVGP 3532

```

```

RESULT 37
A54849
collagen alpha 1(VII) chain precursor - human
N/Alternate names: procollagen alpha 1(VII) chain
C/Species: Homo sapiens (man)
C/Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #text change 15-Sep-2003
C/Accession: A54849; PH0844; S16316; I56328; A30296; I84686
R/Christiano, A.M.; Greenspan, D.S.; Lee, S.; Vitto, J.
J. Biol. Chem. 269, 20256-20262, 1994
A>Title: Cloning of human type VII collagen. Complete primary sequence of the alpha (VII)
A/Reference number: A54849; MUID:94327588; PMID:805117
A/Accession: A54849
A>Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-2944 <CHR>
A/Cross-references: GB:I02870; NID:9987124; PIDN:AAA75438.1; PID:9987125
R/Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A>Title: Molecular cloning and characterization of type VII collagen cDNA.
A/Reference number: PH0844; MUID:92231902; PMID:1567409
A/Accession: PH0844
A/Molecule type: mRNA
A/Residues: 'EPR', 340-475, 'RALSTASHSTLCWRHPCNRGSHWTRAPCEPCNRPASHPAAG', 524-528, 'C',
A/Cross-references: DBJ:D11152; DBJ:D13694; NID:9453698; PIDN:BA02853.1; PID:9453699
A/Experimental source: keratinocyte
A/Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R/Parente, M.G.; Chung, L.C.; Rymmenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A>Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A/Reference number: S16316; MUID:91334380; PMID:1874109
A/Accession: S16316
A/Molecule type: mRNA
A/Residues: 815-892, 'E', 894-1439 <PAR>
A/Cross-references: GB:M6518; GB:S49017; NID:g180914; PIDN:AAA6439.1; PID:g180915
R/Experimental source: keratinocyte
R/Gammon, W.R.; Abernethy, M.L.; Pedilla, K.M.; Priyaanah, P.S.; Cook, M.E.; Wright, J.;
J. Invest. Dermatol. 99, 691-696, 1992
A>Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot
A/Reference number: I56328; MUID:93107742; PMID:1465284
A/Accession: I56328
A>Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RGS>
A/Cross-references: GB:S51236; NID:g262308; PIDN:AA24637.1; PID:g262309
R/Selzler, U.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Ghanville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A>Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A/Reference number: A30296; MUID:89139437; PMID:2537292
A/Accession: A30296
A/Molecule type: protein
A/Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 12032, 'C', 2034-2041, '
A/Note: two reported peptide cannot be reliably located
R/Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A>Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A/Reference number: I48103; MUID:93271985; PMID:8499916
A/Accession: I48103
A>Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 2395-2871, 'S', 2873-2944 <RE2>
A/Cross-references: GB:I06862; NID:g388713; PIDN:AAA8196.1; PID:g388714
R/Christiano, A.M.; Rymmenen, M.; Dittio, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A>Title: Dominant dyotropic epidermolysis bullosa: identification of a Gly --> Ser subs
A/Reference number: A55255; MUID:94224777; PMID:8170945
A/Accession: A55255
A/Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C/Genetics:
A/Gene: GDB:COL7A1; EBR1, EBD1, EB
A/Cross-references: GDB:128750; OMIM:120120
A/Map position: 3p21.3-3p21.3

```



```

Db 680 SPLHSGRFDMWKKSPILRARPRTP-----SEASHRIPESVTLPSNRTSAGTSSSGVNR 734
Qy 106 KRPELLELPDLPLPPSLATGCPAGSEDLTKDRLSTGKLEPVPSPPHIDPELELVP 165
Db 735 KRKKRVSP---IRSEP-----RSPSHMRTSGRLSSSELSPLTPPSSVSSLSISVP- 786
Qy 166 RLSEKEILQNNDRVDREITWVEOISKLKKOQOLEEBAKPPPEPEKVPSPPIESKRS 225
Db 787 -LATSALNPTTFPSSHILT-----OSGEAEKQKPRKKTSAPEAPFSSSTPL- 835
Qy 226 LVQIITYDENRKKAAHRIEGLAPOVELPLYNOPSDTROHYENIKINOARKLILYFK 285
Db 836 -----FPMWTFPSQTERGRNDKAPSELSD----- 861
Qy 286 RRRHARAKOMKOFQCORIDQMLEAKYERLENNPRRAKESKYREYEKOPPEIRKORE 345
Db 862 -RDADKSVEKDK-----SREDREREKENRESEKER-----KKGSEIQSSA 904
Qy 346 L-----OERMQRVGGSGLSMSAAR-SEHEVSEIIDG-----LS 380
Db 905 LYPVGRVSKKEVVEDATSSAKKATGRKSSSHSDGTDITSTVLGDTTAVTKILIKK 964
Qy 381 EOENLEKQMRQLAVIDPMLYDADQRIKFINNGLMADPMKYVDQRYMMMSBOEKETP 440
Db 965 GRGULEKTNLDLGPAPSL-----EKEKTLCLSTPSSSTVK-HSTSSIGMLAQADKLP 1018
Qy 441 REKMOMPRKNGLIASFLEKTVACVLYYLTUKNEYKSLVRRSYRRRQKSOQOQOQ 500
Db 1019 TDK-----RVAASLKKAKAQLC-----KIEKSKSL 1043
Qy 501 QOQOQOQOQOQMPRSSOEKDEKEK-----EKEAKEKEKEPEVE-- 539
Db 1044 -----KOTQOPKAGQGESSETSVTCPRIKAVCRPAVALGKRAVFPDDMFTLSAL 1096
Qy 540 ---NDKEDLKEKTDTDGSEDNDEKAVASKGRKTANSQGRKGRITRSMANEANSEAI- 596
Db 1097 PWBEREKL-----SSMGNDKSSIA--GSEDAEPLAPRIKPIKPTRYAKAPDEPVPK 1147
Qy 597 ---TPQOSAEIASHELNESSHWTEEMETAKKGLIHRMWSALARMGSKTVSOCKNFY 653
Db 1148 KGRSRRCQCGPCGVPEDCVCTNCLDKPRFG--GRN-----IKKQC----- 1188
Qy 654 FNYKRRONLDEILQOHLKMEKENARKKKKAPAAASEAAPPVVEDEMEASGVGN 713
Db 1189 CKMKCCQWLQ--WMPSKAYLOKAKAVKKKKKTS-----EKDOKSESSVKN 1236
Qy 714 EENVEAEALHASGNEVREGESG-----PATVNNSDTESIPSPH-----TEAKD 761
Db 1237 ---VDSQKPTPARBDBPAKSSSEBPPRKPVEKSEEGNVSAFGESEKQATTPASRK 1293
Qy 762 TGQNGKPRPATLGADGPPPPPT--PPRTSRAPLEPTPASEANGAPPPAPSPS-- 816
Db 1294 SSKOVSPALVI-----PPQPPPTGPPRK--EVP-KTTPSEPKKKQPPPESSGEQSK 1345
Qy 817 --APPVVPKKEKEETAAAPVEGE-----EOKPRA----- 847
Db 1346 KVARPSPVAVOKPEKEKEPPVANKQENAGTLNLTSLTNGNSKQKIPAGGVHRIRVDF 1405
Qy 848 ---AEB-----LAVDTGAE-----EPVSECTEEAAE 872
Db 1406 KEDCEAEVWEMGGILITSVITPRVVCFLCASSGHEFYVCQVCCGPFKFLCEE-NE 1464
Qy 873 GPAP-----GKDAEAAEATAE-----GALKKEKE-- 897
Db 1465 RPLDDEJENMCRCCKFCHVGROHQAOTKOLLECNKCNYSYHPECLAGNYPTKPKTKKV 1524
Qy 898 -----GSGGATTAKSSGAPQDSDSA-----TSADAEVDEAB 930
Db 1525 WICTKVCARCKSCGSTITPKGMDAOWSHDFSLCHDCAKLPAKGNFCPLCDKCYDDDDYISK 1584
Qy 931 -----GQDKNRLLSPRPSILTPT-----GDRANASPOKPLDL 963

```

```

Db 1585 MMOCKGCDRWVHSCENILSDEMEILSNLPESVAYTCVNCTERRHPAEMWLAKELQISL 1644
Qy 964 KO-----LKQRAAALPIQVTKVHEPPREDAPTKADAPAPPQNLQDES 1009
Db 1645 KOVITALLINERTTSHLRKRAAKRPDL-----NPEITESLPSSSPEGPPVLTVEYSK 1699
Qy 1010 DAPQPGSSPRGKRSBPAPADKEAFAAEQKLRGDPDPCWTSGLPFVVPREVIKASPHA 1069
Db 1700 QKKQP-----LDLEGYKRMQDQNGTYSVLEFSDDIYKIIQOALINS 1740
Qy 1070 PDPSAFYAPRPHPLPLGLHDTARP-----VLP-----RPTTISNPPPLISS 1111
Db 1741 DG-----GQBEIKKANSMVKSFFRIQMERVPFWSYKKSFWMEBNKVSSNSMGJPN 1791
Qy 1112 AKHSVL-----ERQIGA-ISQMSVOLHAYSEHAKRPVPTVMGLPLMPDKCLA 1162
Db 1792 AVLPSSLDHNAQMOEERENSHTEOPLMKKTIIP---APKPKGDEPDSPTPLPPTPP 1847
Qy 1163 PFSGVQKOBLSPPRQAGPPESL-----GVPTAOBA----- 1192
Db 1848 ILSTDRGREBSP--ELNPPGIEDNRQCALCLTYGDDSDANAGRLVYIGQEWTHVNCAL 1905
Qy 1193 -----SVLRGTALG---SVPGSITKGIPESTRVPSDAITYRGSIT 1230
Db 1906 WSAEVEFEDDQSLKNVHMAVIRGKQLRCEFCQKQGATV--GCCUTSCTSN----- 1953
Qy 1231 HGTPADVLYKGTIR---IIGEDSPRLDGREDSLPKGVYI-----YEG 1272
Db 1954 -----YHFWCGRANKCVFLDDKVVYCOHRD--LIKGEVPENGEVEVRRVAVDEEG 2003
Qy 1273 -----KKSHVLSYEGMSVT-----QCKSEKGRSSSGPP 1301
Db 2004 ISLRKFLNGLEPPENIHMGSMTIDCLGILNDLSDCEBKLPITGYQS----- 2052
Qy 1302 HETAPKRTYDMEGVRGAISASIEGLMGRAIPERHSPHILKEOHNIRGSIITQGIPIR 1361
Db 2053 -----RVWYSTDARRCVYTCKIIVECRPVVEPINSVTEHDENNITAHS-----PT 2100
Qy 1362 SYVAQEDYLAREKALKREBTPPPPPSRDLTAY-----KTQALGPLKLRPAH 1411
Db 2101 SFTSSSKESQNTAEILS-PPSPDRPPHSQSGCYVHVISKVRIRIPSPSPORSQGC 2159
Qy 1412 EGLVATVKAERSIHEIPREELRHTPELPLAPRLKEGSIITQTPTKYDTGASTGSKH 1471
Db 2160 RPL-----PSAG-----SPTTTHIYIVGDL-LSLSLRISIGSRH 2195
Qy 1472 DVRSILGSPGRTPPVPHLDVMADARALERAQYEE-SLKSRRPTGASSGSGSIARGAPVY 1530
Db 2196 STSSL--SPQRS-----KLIRMSPMRTGNTYSRANVSVSFTGTAT-----DLESSAKYVD 2244
Qy 1531 PELGKPROSPLTYYEDHGAIPA-----GHLPKRGSPVYMKREPTPLQSGSLSS 1577
Db 2245 HVLDG-PLNLSSTLSQNTSTSSNLTQRTVTVGNKNKSHLDGSSSEMKOSSASDLVSKSSL 2303
Qy 1578 KASODRKLTSTPRRIASPHSTVPEHHPHIISPFEHLILRGVSDVLYKSHIPLAFDPTSI 1637
Db 2304 KGEKTKVLSKSGE--GSANVVAIPGIP-KLAPQVH--NTTSBELANVKISGSAPEBSV 2357
Qy 1638 PRGIPLDAAAAYVPLRHLAPNPTYPHLYPPYLIRGY-----PD----- 1675
Db 2358 -----SFSKSEAL-----SEPHILH-----LGGQRDRDQHTDSTQASNSPDEDETEV 2399
Qy 1676 -----TALBNQTIINDYITSQO-----MHNTATAMAQRADMLRG----- 1712
Db 2400 KTLKLSGNSNSSLINHEMGSSSRDROKSKCKETPKERHSSKSLFLEQOVTTGEBGN 2459
Qy 1713 LSPP--ESSILANT--AAGPRGIIDLQVPHLPVLVPPPTGPATAMRLAVLPAPO--- 1766
Db 2460 LKPEFMBEVLTPREMGQRPCNNVSSDKIGDGLSMPCVPKAPRPMQVGSASAELOAPRXT 2519
Qy 1767 -----PPSR--HSSBPLS---PG-GPTHLYTPPTTSS 1793
Db 2520 VKVTLTLPKMEBESQKVALKESVTLPLFANRVNISTEPISSAENPDGCPVAQPSPNNTSC 2579

```

QY	1164	FSGVQEBQLSPRGQAPRESIGVPTAOASVYLRGALGSVGGITTKIPIESTRNPDSOI	1223
Db	528	HSGSEESDDEEAPAAAS-----AAQAK-----PALEKOMKASRKTPA-----	569
QY	1224	TYRGSITGTPADVLYKGTITIRIIGEDSPSRILDRGREDSPKGVHYIEGKKHYLSYEGG	1283
Db	570	---SATSITSHCKAGAVTISASLSSP-----ALAKG-----	599
QY	1284	MSVYQCSNEDRSSSGPHETAPKRTYDMMEGRGRAISSASIEGLMGRAIIPERBSPH	1343
Db	600	---RORSVD---SSSESESEGAAP--STRPVQSGSG-----KGLQKALQGVAVP	645
QY	1344	HLKCOHNRIGSITQGISPRSYEAQED--YLREAKTLKREGTP-----PP	1387
Db	646	HTQTKG-----PSVKAQGEDSLSLEDSSSEEDBETPAQATLGLPLQAKANP	695
QY	1388	PPSSDLTEAYVYQALGPLKLRPAHEGLVAYVKEAGRSIHEIPRELRTPELPLAPRLK	1447
Db	696	PTKTPPASASGKAVAPLTKGKPPVYN--STVASRQ-----RSVPAAGKAGAPAT	743
QY	1448	EGSITQIGPLKXPDGASITGSKNDVRLISPGRTFPVNPVLVMDARLYERACTEES	1507
Db	744	Q---AQKPPVA--GTGEDESSSKESDEEBETPAQ---IKPVGKTSQVRAASAPAKSP	795
QY	1508	LK-SRPGTASSGSGSIARGAVIYVELCKPQSPPLYE---DHGAPAGHLP-----	1555
Db	796	KKGAHFGTCKGTGSSATQAP-----QKTEDSDSSSESSSDTMEPAQAKSPVSVN	849
QY	1556	RGSPVYMEPTPLQEG--SLSSSKASODRKLSTPREIAKSPHYVENHPHPISPYEH	1613
Db	850	RNSPPAVAPAPPP---EGQAVNTTKKASGTTAOGSSSESDDEDLIPATQP---STVAL	903
QY	1614	LLRGVGVDLKRSHPLAFDPTS-----IPRGPLDAAVYLPRLHAPPTYPHLYRP	1667
Db	904	RTSVTTTPALSR---ASQPSKSEQSSRMFKKAKAAMAAQSTSSAVETLPMMP----	954
QY	1668	YLIRGVPDTALENRQTIINDYITISQOHNHTATAMQRAAMLRLGLSPRESLSLANTAAG	1727
Db	955	-----PQSATIQPKAT-----NKLGSKLPEKQOALPGVPYKA	986
QY	1728	PRGIITLDSQVHLFVLPVPTGTPATAMDRLAYLPAPOPSSRHSSSPLSPGSPHTLTK	1787
Db	987	PRSESDSEDT-----SSEDEEDAKPOMFK--SAHRLDP--DPSQETVYE	1028
QY	1788	PTTSSSRERDRDRERDREREREKSLITSTTYEHNAPIWPGTQSGSGSSGGGGGS	1847
Db	1029	ETPRESS-----DEMVAQCSLISGYMT-----PLGTVAISOAK-----A	1065
QY	1848	SSRPASHGHAQHSPISPTQDALQORDS-----VLHNTGMKGIITAVEBSKPTVLNST	1901
Db	1066	TPRDSNSILA--SAPATKUNPDGKQKSKSQAUAUTALPKTGRKASGSSTQKPKLKS	1124
QY	1902	STSPVPVPAATFP-----PATHCPLGCTGIDGYITLMEPVILPKEARVAPPE	1949
Db	1125	TSSSPA--PTQILPNSITQRLLEQAMPSEAOVQASVYKVLTELE-----OE	1170
QY	1950	RPRADTGHAFLAKPPASGLEPASPSPKSGSPRLVPPVSGHATITARPANLAPHNASP	2009
Db	1171	RLKA-----TEATRESGKKSQKRLSGDLEGA-----PKN-----KKK	1205
QY	2010	DPPAPASADPHREKTQSKPFSIQLELRLSGYHG--SSVSPGVEPVS-----PVSS	2061
Db	1206	EQVPBRASAVPEKAPMTSKAKS--KLDKGAGGKQKSPQCAKEKPDGELLGITLES	1263
QY	2062	PSLTHDGLGRHLELDKSHLEGRLRQRPQPVVLGGEAHLPHLRPLRESQPSSP	2118
Db	1264	GEQSDPKSKSKKKSLKKK--DKERKEKKKSKSLANDQSA--PLQKKKKKKKSAEP	1318

RESULT 36
A44265
trithorax homolog HTX, version 2 - human

N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C:Species: Homo sapiens (man)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-2003
A:Cross-references: GB:U04284; NID:g184393; PIND:AAA56669.1; PID:g184394
A>Note: sequence extracted from NCBI backbone (NCBIP:117729)
R:Gu, Y., Nakamura, T.; Alder, H.; Prasad, R.; Canaan, O.; Cimino, G.; Croce, C.M.; Cantacciuk, D.C.; Kohler, S.; Cleary, M.L.
A>Title: Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocation
A:Reference number: A44265; MUID:9304666; PMID:1423624
A:Accession: A44265
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-3968 <TKA>
A:Cross-references: GB:U04284; NID:g184393; PIND:AAA56669.1; PID:g184394
A>Note: sequence extracted from NCBI backbone (NCBIP:117729)
A:Reference number: A44264; MUID:9304666; PMID:1423625
A:Accession: A44264
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 63-316, GLINSELEK, 327, Q, 329, VR, 332, DKETPP, 340, T, 342, EDKYVROSPPRR
546, LQIESVP, 255-555, N, 356-359, V, 359-389, A, 390-398 <GDI>
A:Cross-references: GB:U04731; NID:g339921
A>Note: sequence extracted from NCBI backbone (NCBIP:117779)
R:Diabali, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A>Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute
leukemia
A:Reference number: I58112; MUID:93265134; PMID:1303259
A:Accession: I58112
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1317-1700, DD, 1703-1936, H, 1938-2180, S, 2182-2328 <DA>
A:Cross-references: GB:I01986; NID:g307522; PIND:AA9251.1; PID:g553800
R:Marchalek, K.; Grell, U.; Lochner, K.; Nilsson, I.; Stiegler, G.; Zweckbronner, I.; Becr
Br. J. Haematol. 90, 308-320, 1995
A>Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in the
t(11;22)(p15;p11) translocation
A:Reference number: I37165; MUID:95315013; PMID:7794749
A:Accession: I37165
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1212-1603, GTE' <MAR>
A:Cross-references: EMBL:X63604; NID:g897757; PIND:CNA58584.1; PID:g899268
A>Note: submitted to the EMBL/Genbank/DBJ databases by R. Marchalek, 20 December 1994
R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canaan, O.; Salto, H.;
Cancer Res. 54, 2327-2330, 1994
A>Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in
the t(11;22)(p15;p11) translocation
A:Reference number: I38485
A:Accession: I38485
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1251-1486, G', 1488-1538 <RES>
A:Cross-references: EMBL:U04773; NID:g451554; PID:g451555
A:Accession: U04773
A:Gene: GDB:MLL; HTX; ALL-1; HRX
A:Map position: 11q23-11q23
A:Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A>Note: the list of introns is incomplete
C:Superfamily: histone methyltransferase, trithorax protein type
C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F:1434-1456/Region: zinc finger CCHC motif
F:1479-1506/Region: zinc finger CCHC motif
F:1527-1556/Region: zinc finger CCHC motif
F:1569-1596/Region: zinc finger CCHC motif
F:1673-1900/Region: zinc finger CCHC motif
F:1933-1955/Region: zinc finger CCHC motif

Query Match	2.9%	Score 383.5	DB 2	Length 3968
Best Local Similarity	17.9%	Pred. No. 6,4e-06		
Matches 585	Conservative 357	Mismatches 108	Indels 1247	Gaps 143
51	SHLSPGSLIQPRRRPRLS-EPQGNRGSLHLPRSHSYDP----ELGSEMEFTIS			

QY	1272	GRGIIIDUS---	QVPHLEVLVLPRTGPTATMDRLAY-----	1760
Db	2003	SPSSFIIDSCCKDSOSGTAALISPSFDRPHSQTSGCYHVIKSVPRITRBSYSPLONSPG		2062
QY	1761	---LPTAQPF-----	---SRRHSSPLSP-GCPHTLTKPPTTSSS	1794
Db	2063	CRPLPSASPIPTTHEIYTVQDPLLSGLRISIGRRHSTSSLPRLSKLRLMSYVRICSA		2122
QY	1795	ERERDRERDREREREKSIILTSTTVEHAPIWRPGTEQOSSGSGSSGGGSSSRPASH		1854
Db	2123	-----	YRSRVSVSVELGTATPEASAKASDRGCLISSANLGH	2161
QY	1855	-----	SHAHQHSISRTDQAQORSVLHNTGKMIITAVEPSKTV	1897
Db	2162	SAPSSSSQRTVGSGKSTHLDGSSSEVKRCSALDLV-----	---KSLVKGKNTRSS	2213
QY	1898	LRST-----	STSSPRPATPPATPCPLGTLD-GVPTIMEPLVL-KEAPRYAPE-	1949
Db	2214	SKSTDGSASHAYPEIPIKLT--	POVHNAITPELINSKISGAEBETVFFSKSDIVSYQOL	2271
QY	1950	---BRADYHAFIAPKAPPARSGLEBPASSPSKSEBRPLVPPVSGH-ATIARTPAKULAPH		2005
Db	2272	HLRQGRSDRDQHM---DPSQS-VKP--SPNEDGETIKTLFGMGHRPILHEHIGSSSRD		2325
QY	2006	HASPPRPAPASADPPHREKTOQKRFSLQELERSL-----	---GYH-----S	2046
Db	2326	RROKKKSKSKETCKEHSKSYLEBGOVTTGEBGNLKEFADEVILTPFLIQRPCNNVSS		2385
QY	2047	SYPSEGVFVPVSSPSSLTHDPKGLPKHLE-----	ELDKSHLEGELR---PKOPGPVK	2085
Db	2386	EKIGDKVLPISGVPGSGQSTQYEGSSKELEQAPRKCSVKTPLKMBEENSXKTYQESG--		2443
QY	2096	LGGEAAHLPHLRPLUESQPSSSPILQIAPGVKQHQRVVTLAQHISEVITQDYTRHHPOOL		2155
Db	2444	--GSPAHIESVCAPAEVASRSP--GAGPBGVQSP-----	---NNTLSQDPQSNVYQNL	2490
QY	2156	-----SAPLAPLYSF-----	PGASCPVLDRRPSDLYL-PRDHGAPARG---	2186
Db	2491	PEODRNLMI PDGPKQEDGSPFRRYPRRSA-----	---RARSNMFGILTPLYGVRSYGEDI	2544
QY	2197	-----SPHSBGGKSRPEPN-----		2210
Db	2545	PFPYNSUTSKKKGKGSABEQVDCADDLSTSDBDLYYNNFTTIVISSGEBERLASHNLFRE		2604
QY	2211	-----XTSYLGGEGDGEIE-----	---PVSPREGMTFPG	2235
Db	2605	EEQODLPRISSOLDVDGDTEBSTDSTVATSRKSSQIPKRNKNGENENLKIRPDEAGSKE		2664
QY	2236	H-SRSAYVPLRYDGEQTEP-----	SRNGKSPGNTSOPPAFFSKL-----TE	2277
Db	2665	HVIXSAV-----	GHKNEPKLNDCHSVSRV--KAQGDQSL- EADLSLSSRRVHTSTP	2714
QY	2278	SNSMAVVKSKQOEINKLMLTHNRNBEVYNISQGTGFI F--	---NMPAITGMLTYTSQAQOE	2334
Db	2715	SDKNLIDTYNAELLSKSDSDNNNSDDCGNLPESDINDFLKATPSKQALGESBESSSEL		2774
QY	2335	HASTMGLAEIIRKALMGKYDOWEESPILSANAFNPILNASAS--		2383
Db	2775	TLGEBLGDLS--NRREDIGLFEVFSQQLP---	ATEPVDSSVSSSISABEQFLPHELPSD	2829
QY	2384	-----	---TAADGRSDHTLTSPG-----	2399
Db	2830	LSVLITRSPVPSQNPBSRLAVISDSGEKRVITTEKSVASSBEGDPLMLBPGVDPAPEGHMT		2883
QY	2399	-----	---GGGKA---KYSGRPSSRKAKSP-----	2421
Db	2890	PDHFIQGMADHDHISPPCGSVYEQHGANSODLTRNSGTFGLQVPSVPLVPVONOKTYVSS		2949
QY	2442	ASGRPRPVSYSVHSGDCNRRTPLTNR-----	---VWEDRPSAG-----STEPY	2464
Db	2950	TDSPEBSQISMAAYQTTPEPHLKPAIEKXILVNNQNNQPLVLYQTLPENGVTKQIKQLTSFVSS		3009
QY	2465	NPLIRLQAGVW-----	---ASPPPGI--DAGSGPLAPPH	2496

[illegible]

R. Rubinstein, A.L.: Broadwater, A.H.: Lowrey, K.: Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Fox genes: pollen-specific genes with extensin-like domains.

A:Reference number: S49915

A:Accession: S49915

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1188 <RUB>

A:Cross-references: EMBL:23465; NID:G600117; PIDN:CAA84230.1; PID:G600118

Query Match 2.9%; Score 389.5; DB 2; Length 1188;

Best Local Similarity 18.6%; Pred. No. 1.1e-06;

Matches 292; Conservative 104; Mismatches 369; Indels 807; Gaps 64;

```

QY 733 KGCSPRATVNNSSDTEISIP-----SPHTAAADTGONGPKP---PATLGADGPPRG 781
DB 374 RNDICGLALPRAOKTALQCAFLVLRPVDCSHVCAGYPTPGCGSSPVKPKPAASAPMS 433
QY 782 PPTPRRTSRAPRPPASATGATPTPPRPAP--PSPSAPRPVVKKEKEEETAAPVEEG 840
DB 434 PHTPP-----DVSPRLPEPSFVPAPAPMWTPTHSPP-----A 467
QY 841 EEOKPPRAEELAVDTCKAEKPVKSECTEAEKSPRAKDAEAATAEGALKAEKKGGS 900
DB 468 DDVVPPTP-----PVPKSPRA-----484
QY 901 GRATTAKSSGAPODSSSATCSADENDEAEGDKNRLLSPRSL--LTPTGDRANASPO 958
DB 485 -----TSPSPQVQBPAAAS-----TPPSELVTKLSP-----1015
QY 959 KPLDLQKQKORAAIPIQVTKNPRREDAAFTKAP---PAPRPQNLQRESDAPOQ 1015
DB 511 APV-----GSPRPVKT--SPAPIGSPSPRPVSVSPRPVKSPP---PPAP 555
QY 1016 GSSPRGKSSRPAPADKFAFAEAOKLPGDPCMTSGLPFPVPREVIKASPHAPDPSPAF 1075
DB 556 VGSPPRPKSPPPRA-----PVASPP-----PVKSPRPPTLVAS--PPRPVK 596
QY 1076 STAPRPHPLPLGLHTARPLPPRPPTISNRPPLISSAKHPSVLERQIGALISQMSVOLHV 1135
DB 597 SPPPRA-----PVASPPRPVKSPPRPPTPVASPP-----624
QY 1136 PVSENAKAPVPTMGLPLPMDPKLAPFGVQKQEQDSPRGQAPRESLGVPTAQAESVYL 1195
DB 625 -----PPAPVASSPP---PMKSP-----PP-----642
QY 1196 RGTALGSVPGSITKGIPTSTRVPSDAITVRSITHTPADVLKGTITRIIGEDSPSRL 1255
DB 643 -----TPV-----645
QY 1256 DRCGRDLSLPGKHVITGCKGHVLSYEGGMSVTQCSKEDGSSGPRPHETAPKRTYDME 1315
DB 646 -----SSPP-----649
QY 1316 GRVGRATISSATIGLMGRAIPRRHNSPHNLKEQHNRGSIITQIPRSYVAQEDYLRRRA 1375
DB 650 -----PPPK-----653
QY 1376 KLIKREGTTPPPPSRDL--TEAYKTOALGRLKRAHEGLVATVXKAGSIHEIREEL 1433
DB 654 -----SPPPRPASTPPPEYPTP--PTSVKSSP-----PPKKS 686
QY 1434 RHTPELPLAPRLKESITOGTPLKDTGASTGSKKHNVRLIGSBGRTFPPVHPLDVA 1493
DB 687 LPPTPLIPSPPP--QEKPTPSTPSKSPS-----SPKSPSP-----721
QY 1494 ADABALERACVESSLKRPETASSSGGSIARGAIVIVPELIGKRGOSLTEDHGAAPAGH 1553
DB 722 -----KPPVSSPPTPKSSPPPAVSSPPPTVSSPAPLAPVS---SPPSVKS 766
QY 1554 LPRGSPVTMKAPPTPLQEGSLSSSKASODKRLTSTRELIKSPHSTVPEHHNPISPYEH 1613
DB 767 SPBPALSSPPAPVQK-----SSPPPVVSSPPPAKSSP--PLAP---806

```

```

QY 1614 LARGVGVULYKSHIPLADPTSIKGIPLDAAAYVLEPRHLAPNTYHLY---PPYL 1670
DB 807 -----VSS-----PPVKEKTS--PPAPLSSPP-----LAPKSPPHVVVSSPPVP 846
QY 1671 KGYPTPALLENQTIINDYITSGQMHNATAMQAQADMLKGLSPRESSLALNYAAGPRG 1730
DB 847 KSSPPAPVSSPP-----LTKRAPAPVHSSPP-----876
QY 1731 IIDLSQVPHLPVLPPTPTPATMDRLAYLTPAQPFSSRHSSSPSLFGCGPTHLTPTT 1790
DB 877 -----VKPSTPPAPPTVISP-----DSBPk-----SSP--PPTVSLPPTIV 912
QY 1791 TSSSERERDRDRDREREKKSILSTTTVTHAPITWRCTGSSSGSSGCGGSSSR 1850
DB 913 KSSPP-----917
QY 1851 PASHSHAHQHSPISPRTDOLQORPEVLNHTGKGIITVAPSKPTVLKSTS---TSSP 1906
DB 918 PAMVS-----SPMTPKSSP-----PPVV-----VSSPPPTVKSPPPAVSSPP 957
QY 1907 VBPATFPATPCPLGTLIDGVYPTLMEVULLPKKAPVAPRBERPRADTGHAFLAKPPAR 1966
DB 958 ATPKSSPPRA-----PVNLPPREKSSPPTPVs-----SPPA- 991
QY 1967 SGLERASSPSKSGSERPLVPVSGHATITARTAKNLAEPHNSAPRPARASADPHREKT 2026
DB 992 -----PKSSP-----PPAMSSPP-----PPVKSPPPPAPVSSPPPP---V 1025
QY 2027 OSKPFSTIQLELRLSLGTHGSSVSPGEGVPSVPSSPSTLTHDKGLPGLHELDKSHLEGEL 2086
DB 1026 KSP-----PPAVSSP-----1037
QY 2087 RPKQRPVVLGGEAHLPHLRPLPSOSPSPLLOTAPGVKGRVVLTAQHISEVITOD 2146
DB 1038 -----PPVKSPPPPAPVSSPPPPVKSPPPPADISSPPPVK-----1074
QY 2147 YTRHHPQQLSAPRLAPVLPFGASCPVLDLRRRPSDLVLRPDHGAPARGSPHSEGGKS 2206
DB 1075 -----SPPPAVSSPP-----PPVKSPPPPAPVSSPPPPKSPPPAPVS---S 1116
QY 2207 PERNKTSVLGGEDGIEPVSPGEMTEPGHSRS---AVYPLLYRDEQTEPSRMGSKSPG 2263
DB 1117 PPP-----APVKKP--SLPPAPVSSPPPVITPAPPKKEQSLP-----PP 1155
QY 2264 NTSQPPAPFSKL 2275
DB 1156 AESQPPPSFNDI 1167

RESULT 34
A48205
A11-1 protein +GTE form - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 22-Jun-2003
C/Accession: A48205; B48205
R/Name, Q, J, Alter, H, Nelson, K.K.; Chatterjee, D.; Gu, Y.; Nakamura, T.; Canaan, E.; Crc
Proc. Natl. Acad. Sci. U.S.A. 90, 6350-6354, 1993
A/Title: Analysis of the murine A11-1 gene reveals conserved domains with human A11-1 and
A/Reference number: A48205; WUID:93317679; PMID:8327517
A/Accession: A48205
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-3869 <HAR>
A/Cross-references: GB:LI7069
A/Accession: B48205
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1505,1509-3869 <HA2>
A/Cross-references: GB:LI7069
C/Genetics:
A/Gene: A11-1
C/Superfamily: histone methyltransferase, trithorax protein type

```


[illegible]

Db	1866	LSP-----SSIKVEGIEDFESPS--	-MSSVNLNFDTQKLDNDCCSVNTALTITDTTGDGEG	1937		
Qy	1468	SKKDVRSLJSPGRTFPPVHPLVDVMDARALERACYEESLKSRRGTASSGSGS	IARGAP	1527		
Db	1938	NADND--SATIATETTSASSAPNEGILTYAAMAMSEYEDRLSS--GLVSPA-----		1984		
Qy	1528	VIVELGKPROSPPLYEEDHGA	PPAGHLP	RGSPV	TWREBPTPLDQGLSSSKASQDR---	1583
Db	1985	-----PSFYSKSEYDNEGT--	-VDYBETSSLIADPCSPSPG-	ASGSAGKSGSDSDRGQK		2033
Qy	1584	-----KLITSPREIKSPHISTYRENNHPISTYEHLLAGVSGVDLYRHHIPLAPPTSI	PR	1639		
Db	2034	RFTQTMNLQKVLKSCFND-----		YKPTMLECCVLGNDI		2069
Qy	1640	GIPLDAAAYY-----	-LPHILAPNPT-YPHLYPPLYIRGVPDPAALENQTI	1685		
Db	2070	GLPKRVQWYQNAARAEKSKLSMAGHFGINGQSYSGPRTECTLGKITYSARISVYDHI		2129		
Qy	1686	-----INDYTSQ-----QMHNTATA--	-MAQR-ADMILGSLPRESSIALNYAAGR	1729		
Db	2130	FSQGHISKVKOTISQDDEKEYEYDPDPAIVRQDMAQDELDRI-----	-KKANEVLGLAAQOQ	2184		
Qy	1730	GIIDLS--QVPHLVLVLPPTGTATAMDRLAYUPLTAPQPPSSKHSSPLSPG--	-GPTH	1784		
Db	2185	GMFNPTPLQALNPLTAPYALOGIPPLVPLGL-----	-NSPSLPGFTPENTA	2229		
Qy	1785	LTKPTTSSSERDRDRERDRORERKESILSTTTVEHAPIMRPTGEOSSGSGSGCG	1844			
Db	2230	LTSRKPNIMG-----	-LPSTT-----VSPGL-PTSGLPNKSSA	2262		
Qy	1845	GGSSSRPASHSHAHQHPISPRTDALQORPSVLYANTGMKGIITVAEBSKPYTLRSTGS	1904			
Db	2263	SLSSPTPAQATMA--MGPPQRPQOQOQOQOQOY-----	-QQPRP	2299		
Qy	1905	SPVPRATFPRATNCPLRGILIDGYPTPLMEVULLPKAPRYARERPRADTGAFLAKP	1964			
Db	2300	---PALQPPPTPOLPL-----	-QOQOORQDQSEKYEKEXKHKGKGP-LPVPK	2344		
Qy	1965	ARSGLEPASPSSKSEBRPLVPRVSGHATTA	TPRAKYALARHNASRDPAPASDPHRE	2024		
Db	2345	KEKEATPATATATISALPTM-----	-EYAVDPAQOQA-----LOALITSDPTAL	2388		
Qy	2025	KTGQKPFESIOELBLSLGHGSSYSRGEVPRVSPSPSLTHNDKGLPKHLELDKSHLEG	2084			
Db	2389	LT-----	-SGPLFYGVGFSYUAPQI-----	2409		
Qy	2085	ELRKPQGPVYKLGGEANHLPHLRPL-----	-PESQDSSPPLIQTAPG--VKGHQRYVT	2134		
Db	2410	-----EGALQSG-----	-YLPWYGMEGILFYSPSPALSQALMGLSPGILLQOYOOQS	2456		
Qy	2135	LAOHISVITPODYTRHNPQLSAPLRYAPLCSFPGASCPVLDLRRPSDLYLPRPHGARA	2194			
Db	2457	LOEALIQOQOQKYOQOQPKASQTPVP-----	-DGA-----PSPDXD-PA	2493		
Qy	2195	RGSPHSESGKRSPB-----	-PNKTSVYLGGEGDI-BEVSBPB-----GMT	2232		
Db	2494	KESPKEQGNKNTREVSRPLPKLPEBEBEASKSADSLYDPFIYPKYUUKVCCKQAGPS		2553		
Qy	2233	EPGHSNATVPLLYRDEQTEPSRM-----	-GSKPAGNTSQPAPFSKL	2275		
Db	2554	DEEABRSHLSLCPFGOSVNLQEMVILHVPYTGCGGSGGSGGSGGSGGSGGSYCHLA	2613			
Qy	2276	TESMSAMVSKKQKEINKUKLNTHN-----	-RNEBYNISQPTGTEIMMPALITGILMTYS	2329		
Db	2614	CEBALGCEBALSHLESALHKKHTTTAAANNAEHPSLLPHSACFPDPFSTA	STASQGAHS	2673		
Qy	2330	-----QAVOEHASTNNGLBALIRKALMGKYDQWEESSPPLSANAFLNLSASL	2377			
Db	2674	NDSPPPPSAABPSASBHSKRKMPQVGRAS-----	-AAKPP-----SFPPLSSSTIV	2721		
Qy	2378	PA-----	-AMPTAADGRSDHTLTVSPGGGKAKVSG	2407		
Db	2722	TSSSCSTSGVQPMPTDYESEEDTDLSSKSDPAPAPVEG	2761			

Db 843 GGNHAFNFWTPRPPL-----PPASIGAAVQPKMESGT-----PAGPPNVLPLSM 890
Qy 1217 VPSDSAITYRGSIITHTPADVLKGTITRLLIGDSSRLDRGDESLPKGHVIEGKGK 1276
Db 891 APPLSL-----GLRGHGA-----QTEPTKVE-----VKPVP----- 917
Qy 1277 VLSYEGMSTVQCKEDGRSSGPPHETAPKRTYDMERVGRAISSAIEGLMRAIP 1336
Db 918 -----ASPPHKKVSAVLVSPQW-----KALACVSAEGVTEEPA 952
Qy 1337 PERSHPHLKEQHNGIRGSIQTGIPRSYVEAQEDYLREAKLIREGTPPPPSRLDTEA 1396
Db 953 SEHLKP-----ETQETPRPEKPLRYTKAVPT-----RQST-V 985
Qy 1397 YKTQALGPKLKPAGHGLVATVKEAGRSIHIEPRELHTPELPLAPRLKEGSIQTGTP 1456
Db 986 PKLPAVHPARLR-----KLSFLPTPR-----TQGE 1011
Qy 1457 LKQDTGASTTGSKKHIVRSII-----GSPQRTPPVPHPLDVADARALEACYEESLSKR 1511
Db 1012 DVQAFISETIGIASLSSILEQPEKSEAKKECPAPADSLA----- 1054
Qy 1512 PGTAASSGGSIGAGAFIVPELGK-----RQSPITYEDHGAPFAGHLPRGSPVTMRREPTPR 1568
Db 1055 -----VGNSSG-----VDIFQEKRPDLRLQAP-----ELANVAGLTP---PAT---PPHQ 1093
Qy 1569 LQEGSLSSSKASQDRKLTSTPREIANSPHSTVEHHHPHPISEYHLLRGVGVND----- 1622
Db 1094 LMKPLAVALSIKAKAPKSTAQEGTLKPEGVTEAKHPAAV-----RLQEGVHGSRAVHGS 1149
Qy 1623 -----LYRSHIIPAPPTSPRGIPDLDAAYVILPR-----LAPNTYHILVPPYL 1669
Db 1150 GDHDYCVRSRTP-----PKKMPALVIVPVGSRMVKRQDITIKPVLSLGAAP---PP-- 1200
Qy 1670 IRGYPTAALENQTIINDYITSCQMHNTATMAQADMLRGISPRESSIALNVAAGR 1729
Db 1201 -----PCIAA--SEPL--DHRTSEQADPSAPCLAPES-----LLSP-EASPCN----- 1241
Qy 1730 GIILDSOVPLPVLVPTPTPTATAMD--RLAYLPTAP-----QPPSSRHSSPLSPGCP 1782
Db 1242 -----DMNTRTPPEPSAKQSRMRCYKACRSAPSSQGWGRGRSRSVSSGSR 1292
Qy 1783 THLTCTTTSSSRERDRDRERDRERREKSIITSTTVHAPITMRPGTQSSG----- 1836
Db 1293 TSEBASSSSSSSSSSSRSLSPPHKR-----WRSSCSSSGRRRCS 1336
Qy 1837 SSGSGGGGGSSSRPASHAHQHSPISPRTODALOORPS 1876
Db 1337 SSSSSSSSSSSSSSSSSSRSRSP-SFRRSDRRRKYIS 1375

RESULT 31

A41948
alpha-fetoprotein enhancer-binding protein - human
N/Alternate names: ATBFI protein
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C/Accession: A41948
R/Morinaga, T.; Yasuda, H.; Hashimoto, T.; Higashio, K.; Tamaki, T.
Mol. Cell. Biol. 11, 6041-6049, 1991
A/Title: A human alpha-fetoprotein enhancer-binding protein, ATBFI, contains four homeod
A/Reference number: A41948; MIM:204933; PMID:1719379
A/Accession: A41948
A/Molecule type: mRNA
A/Residues: 1-2783 <MOR>
A/Cross-references: GB:010250; GB:D90395; NID:9219429; PIDN:BA01095.1; PID:9219430
A/Note: sequence extracted from NCBI backbone (NCBI:66271, NCBI:66276)
C/Genetics:
A/Genes: GDB:ATBFI
A/Cross-references: GDB:392090; OMIM:104155
A/Map position: 16q22.3-16q23.1
C/Superfamily: alpha-fetoprotein enhancer-binding protein; homeobox homology

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
F:72-94/Region: zinc finger CCHH motif
F:128-150/Region: zinc finger CCHH motif
F:176-198/Region: zinc finger CCHH motif
F:311-332/Region: zinc finger CCHH motif
F:340-361/Region: zinc finger CCHH motif
F:448-471/Region: zinc finger CCHH motif
F:489-509/Region: zinc finger CCHH motif
F:517-538/Region: zinc finger CCHH motif
F:633-655/Region: zinc finger CCHH motif
F:684-706/Region: zinc finger CCHH motif
F:719-773/Region: serine/threonine-rich
F:809-958/Region: glutamine-rich
F:1071-1092/Region: zinc finger CCHH motif
F:1117-1211/Region: proline-rich
F:1232-1288/Domain: homeobox homology <HOX1>
F:1329-1385/Domain: homeobox homology <HOX2>
F:1416-1437/Region: zinc finger CCHH motif
F:1618-1638/Region: zinc finger CCHH motif
F:1728-1784/Domain: homeobox homology <HOX3>
F:1799-1820/Region: zinc finger CCHH motif
F:2033-2089/Domain: homeobox homology <HOX4>
F:2112-2134/Region: zinc finger CCHH motif
F:2545-2566/Region: zinc finger CCHH motif
F:2585-2607/Region: glycine-rich
F:2611-2633/Region: zinc finger CCHH motif
F:2650-2737/Region: serine/threonine-rich

Query Match 3.0%; Score 394; DB 1; Length 2783;
Best Local Similarity 18.2%; Pred. No. 1.9e-06;
Matches 510; Conservative 273; Mismatches 953; Indels 1064; Gaps 115;

Qy 99 EMFTSKPRRLLELDPLRPSPLATGQPSGSELT-----KDRSLTGLEVPSPS-- 152
Db 535 ETSHELSEADIQQLGGLANGDLAMGDPPLAEHTIIVEDKEESDLEDKQSGTGS 594
Qy 153 -----PPTDPELVPR-----LSKEELIQNDVDRREITWV 186
Db 595 DSGSVQEDSGSEKRLAPRKGNFTMEKFLDPSRYKCTVCKESTQ-----KILLY 648
Qy 187 E-QQISLKKQCOOLEEAAKPEPEKPVSPPIESKHSVLVOIYDENKKAAEAHRL 245
Db 649 HYNVSVHHLKRLALDESATGQPEPTSPDNKPFKNTCNVA----- 690
Qy 246 EGIQVVELPLNQSDTQYHENIKINQAMRKLLILYFKRRNHAKQKQKQRYDL 305
Db 691 -----YSQ-SSTLEIHMRSVYLHQ----- 708
Qy 306 MEALKKVERIENPRRAKESVRYREYKEQFEIRKQRELOERMQRGSGLSMSA 365
Db 709 -KRAAKLEAAGSSNGTNGSSSI-----SLSSSTSPVSTSGSNTFTTIS 752
Qy 366 ARSEHEVSEIIDLSEQNELKQKQRLAV-IPMLYDADQOQRIKFTNMGLMADPMKVYK 424
Db 753 NPSS-----AGIASSNLISQVPTESVGMP-----LGNPICA-- 785
Qy 425 DROVMNMSEQK-KETFRKFTMOHPKNGILASFLEKTYAECVLYYLTKNENKSLV 483
Db 786 -----IASPSEKKNRKLAD-----MIS-- 807
Qy 484 RSYRRRGSGQQQQQQQQQQQQQQQQQQPPRSGQEEKD--EKEKEKAEKEEK-- 535
Db 808 -----RQQQQQQQQQQQQQQQQQQQQQTAAQAAQVQAHQQLQQAALQSLRPPL 862
Qy 536 -PEVNDKEDLLLEKTD-----DTSGEDNDEKAVASKGRKTVANSQGRKRTIR 584
Db 863 LPIFPMTTEILLQQQQHLFPFYIPSAIFQNPVSLPVTSGALTLTGTG--PGLLD 920
Qy 585 SWAN---EANSERAITPQSAELAS-----MEINSSWTEEME 621
Db 921 LKAQVQVPOOSHQOILPQOQONQLSIAQSHSALLQPSQHEKKNKLVIEKEKESEQRERD 980
Qy 622 TAKKGLLEGHGNMSALRANVGSKTYSQCKNFYFNKKQQLDEILLQOHLKMEKENARR 681

```

QY 1467 GSKKHIVRSILGSPRTFPPVHP-----LDVMAADARALERACYSBESIKSRPGTAS 1516
Db 900 -----IGPPG---PGENECTPGHRSRPGIPGVKQGRGTGAGAGEODKKNPBPSE 946
QY 1517 SSGGSIARAGAPVIVPELQKPRQSPLETHDQAPAGHLR- RGSPTVTRREPTPLQEGSL 1574
Db 947 ISHVIQKGBRGLKGFAGNPGQEK-----GNRGVCPMPGLKGLRPGAPGPPGR---GDL 999
QY 1575 SSSKASDRKLTSTPRRI-----AKSPHSTV-PEHHPRIPSPYHLRQVSGVULYR 1625
Db 1000 GSTGNPEPGLLIGPSMGMMGMPGSKGKGTLLGFPGRAGRPGLPGLIHGIGDGEGBGYS 1059
QY 1626 SHI-PLAFDPSTIPRGIPLDAAAYVL-----PRLHAPNPTPYPLLYRGPYDTAAL 1679
Db 1060 EGRPRGPGPTGPB-GLRPGMKKGEMGQRPGRHL--GAPGPGAR-----GSPGSPGL 1111
QY 1680 ENRQTLINDYITSQOMHNTATAMORADMLRGLSPRESSIALNVAAGPRGIIDLQVPH 1739
Db 1112 PKG-----PQPHQDLGFKGIRK 1128
QY 1740 LPLVLP---TGTPTATMDRLAYLPTAROPSSSRSSPSLSPGPTHLTKPTTSSSE 1795
Db 1129 L--LGPPGIRGPPGLPG-----PGSPGPMGIR----- 1154
QY 1796 RERDRERDRDREREKSLTSTTTVEHAPIMRPGTESSGSSGSGGGG-----GSSSR 1850
Db 1155 --GDQGRDGLRPGAGEK---GETGLRAB--PGRGNPDAQAKDRGAPGPPGLR 1205
QY 1851 PASHSHAHQSPISPTOTALQORPVLANTGMKGIIITAVEPSKPYLRST--STSSPV 1908
Db 1206 KGMAGDA-----GPRGPTGEGPR-----GPPGLPGAILPGQ-TGNRGPGRSGPGA 1252
QY 1909 PAATTPPATCPLUGTLDGYPTLMERYLLPKEARVYARBERPRADTGHAFKAPRARG 1968
Db 1253 PCPPGPPGSHV-IG--IKGDKGSMGHG--PKGPPGTA-----GDMG--PGRIG 1295
QY 1969 LEPASPSKSGSEPRPLVPVSGHATARTPAKNLARHNASDPDRAPASADPHREKTQS 2028
Db 1296 -----APGTGLRPGRGPD-----GP 1311
QY 2029 KPFSIOLELRSLGYHSSSYSPGEBVPVSPSSPLTHDKLRPKHLELKSHEGLR 2088
Db 1312 QGPPGVKGEKNGPGLG-SIGPPG--PIGKPPGVAGDPGLTKIIS-----LP 1357
QY 2089 KQRPVVLGGEAAHLPHLRPLPSQSPSSPLQTAQVKGHQRVVTLAQHISEVITODYT 2148
Db 1358 GSPGPPGTPGE-----PQMGSE----- 1374
QY 2149 RHHPQOLASPLPVLVSPGASCVLDLRPRPSDLYLRPDHGAAR-----GSP 2198
Db 1375 -----PGD---PG-----PQNLG-PGGRKRGKGGKGRTPGPRAGEK 1408
QY 2199 HSEGGKSPRPNTSVLG-----GEDGLEVPSPGEMTEPGHSRAVVP----- 2243
Db 1409 GNGSGEPBPASDGLRGLKGRKSGSPATYTRTFVTRHSQTLAISCEBGTIVPLV 1468
QY 2244 ----LYRDEQ-----TEPSRMGSKSPGNTSOPRAPFSKLTESNASAVKSKOEINKLN 2295
Db 1469 SGFSFLFVQGNQRAHGODLGTLSCLQRTTMTPLFCN-----VNDVCN 1512
QY 2296 THHRNEPVYNIQPGTEIFNMRAITGTGMLTYRSQ-----AVQEHASTMNGLEAIT 2346
Db 1513 FARRNDISYVLTSTPALMPPMMAPIITGALRPYISRCTVCEGPAIALAHSQTT----- 1565
QY 2347 RKALMGKYDQWESPPLSANAFLNPLNASLPAAMP---ITAADGRSDHTLTPSGGGGA 2403
Db 1566 -----DLP-----CHGWLSTLWKGBFIMFTAGSGGTG 1595
QY 2404 KVSGRPSRRAKSPAPGLASGDRPPSVSVHSGDCN 2440
Db 1596 QALASPOS-----CLEEFRASPLFCHGRGTGN 1623

```

```

RESULT 30
T00273
hypothetical protein KIAA0595 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C/Accession: T00273
R/Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Korani, H.; Nomura, N.; Ohara, O.
DNA Res. 5: 31-39, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A/Reference number: Z14086; MUID:98290545; PMID:9628581
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1520 <NAG>
A/Cross-reference: EMBL:AB011167; NID:g3043713; PIDD:BA025521.1; PID:g3043714
A/Experimental source: Brain
C/Genetics:
A/Note: KIAA0595

```

Query Match 3.0%; Score 395; DB 2; Length 1520;

Best Local Similarity 22.4%; Pred. No. 8.9e-07; Matches 331; Conservative 148; Mismatches 515; Indels 486; Gaps 76;

```

QY 528 EAEKEEKPEVE--NDKEDL--LKEKTDTS-----GEDNDEKAVASKGRKTANSQ 575
Db 251 ESETEAVPKVTLCSKEGSLNSSEKLDACILKREVEVEPVYKPEPPQNPANAAGSQ 310
QY 576 GRRKRITRSMANEANSEAITPQASALASMEINSSRWTEEMETAKKGLLEHGRWS 635
Db 311 RARGRKKSKSQ-----PAACVEGYARRLSSSR-----GQS-- 343
QY 636 AIAKMSKTVSCQKNFYRYKKRQNDLQCHLKMKEKRNARRKKKAPAAASEEA 695
Db 344 ----TVGTEVTSQ-----VDNLQKQPEELQKSGPLQCGKGRARAWAA 386
QY 696 FPPVDEEMEAAGVSGNEEWEAEALHAGSGNEVREGCSGPATV--NNSSDT--ESIPS 753
Db 387 -----ALENSPKULE-----RSAGSSAKK--GPLDIYPLADITQINPI 426
QY 754 PHTAAKDTQNGKRPATVAGDPPGP-----PTP-----RTSRAPIETPASEA 802
Db 427 PTHLSLVDSAQASBMPDVSEADPTAVGVLAVGVPDGLVLDASTSSSEVLEPLAEV 486
QY 803 TGAPTPPAPPSAPPPVYKKEKEETAAAPVEEGEOKPPAAELAVDTCKAEPV 862
Db 487 L--INEVLADSAAVPAVVIDN-----LPPVD--AVPSGAPVDIAL--VDPV 530
QY 863 KSECT-----FEAEKGPAGKDAEAAEATA-----EGALKAEK 895
Db 531 PNDLTPVDVPLVKSRLPDRKGAVALGSAARQLVSESLDPPKTIIEVKEVVDLSK 590
QY 896 KEGSGRATTAKSGAPQDSASATCSADEVDEAGDKNRLSPBSLLTPTGDPANA 955
Db 591 IESGTS-ATTHEA--RRRLPLSEYRRRROQOAEFER--SPQ-----PTG--KMP 637
QY 956 SPQKPLDKOLKQRAAIPPIQYTV-----HEPRRDA-APTKAPAPAPP----- 1001
Db 638 LPEITPGLADIP--CLVTPPAKAKTALQRPETPLEICLVPGSPSPASPSPEPVSKPV 695
QY 1002 ----PQNLQPSDAPQOGSSPRGKSRSPAPPAUKKAAABAQULPGDPCCMTSGLPFV- 1057
Db 696 ASSPTQVPSQEMPLARPSBPVQSVSPAVT-----PQMSAALPPDAG 740
QY 1058 -----PREVIKASPHADPPSAFVSAP-PGHPLPLGLHDTARVLPREPPI 1102
Db 741 GLGMPSPLEPPPIQPSLPLSMGPVLDPPT-HYALPLSPKCPVPHVSPSGYPCLPPTPV 799
QY 1103 SNPPPLISSAKHPSVLERQIGALSQGNVOLHVPYSEHAKAPVGP-----VTMGLPLMDP 1158
Db 800 ----PLVSGTP-----GAYA--VPPTCVPMAP-PPAPVSPYSSTCTYG-PLQMG 842
QY 1159 -KTLAP-SGVKQQLSPRQAGPPESLGVPTAQEASVLKGTALGSPGSGITKGIPISTR 1216

```

RESULT 29
 CCHN3B
 Collagen alpha 3(IV) chain precursor, long splice form - human
 N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
 C/Species: Homo sapiens (man)
 C/Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
 C/Accession: A54763; A43928; A44043; A45971; A39786
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Trygvaason, K.; Reeder, S.T.
 J. Biol. Chem. 269, 23013-23017, 1994
 A>Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
 A/Reference number: A54763; MUID:94364994; PMID:8083201
 A/Accession: A54763
 A/Molecule type: mRNA
 A/Residues: 1-1670 <MAR>
 A/Cross-references: GB:M92993; NID:G577563; PID:G577564
 A/Experimental source: Kidney
 R/Thirner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
 J. Clin. Invest. 89, 592-601, 1992
 A>Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al
 A/Reference number: A43928; MUID:92147878; PMID:1737849
 A/Accession: A43928
 A/Molecule type: mRNA
 A/Residues: 1331-1524, 1', 1526-1670 <TUR>
 A/Cross-references: GB:M81379
 A/Experimental source: Kidney
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 267, 19780-19784, 1992
 A>Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpastur
 A/Reference number: A44043; MUID:93015826; PMID:1400291
 A/Accession: A44043
 A/Molecule type: DNA; mRNA
 A/Residues: 1386-1670 <CUI>
 A/Cross-references: GB:M92993; NID:G177895; PID:AAA21610.1; PID:G177896
 A/Note: sequence extracted from NCBI backbone (NCBI:P115597)
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
 J. Biol. Chem. 269, 17358, 1994
 A/Reference number: A44738; MUID:94274734; PMID:8006044
 A/Contents: annotation; exaltum; correction to intronic sequence in A44043
 R/Bernal, D.; Quinones, S.; Saus, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A>Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A/Reference number: A45971; MUID:93280184; PMID:8505332
 A/Accession: A45971
 A/Molecule type: mRNA
 A/Residues: 1427-1444 <BER>
 A/Note: sequence extracted from NCBI backbone (NCBI:P133363); sequence incorrectly ident
 R/Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reeder, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A>Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
 A/Reference number: A39786; MUID:91355570; PMID:1882840
 A/Accession: A39786
 A/Molecule type: mRNA
 A/Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A/Cross-references: GB:555790; NID:9234418; PID:AA19637.1; PID:9234419
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C/Genetic:
 A/Gene: GDB:COL4A3
 A/Cross-references: GDB:128351; OMIM:120070
 A/Map position: 2q36-2q37
 A/Introns: 1385/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete
 A/Note: The alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C/Description: minor structural component of extracellular basement membrane in kidney
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F:1-28/Domain: signal sequence #status predicted <SIG>

F:25-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
 F:25-42/Domain: amino-terminal nonhelical, NH1 <NH1>
 F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predi
 F:253/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
 Query Match 3.0%; Score 397; DB 1; Length 1670;
 Best Local Similarity 21.1%; Pred. No. 8,4e-07;
 Matches 388; Conservative 121; Mismatches 624; Indels 704; Gaps 91;
 QY 752 PSHTFAKKTQNG-PKPRATIGADGP-----PPRPPTPRRTSR-----AATP----- 796
 D 343 PTEYTDYQKGBGTGGPPGPRGARGQSPGPGVPSG-SSRGLRGAGCWPLIG 401
 QY 797 -----TPSEATGATPPRPPSPSPAP-----PPVVKKEKEBETAAPVEEGEEOKPP 846
 D 402 SKGERGAPGKANGTOSPCACSPGLPSGPP-----GPGDIDGFRKGP 448
 QY 847 AAEEL-----AVDTGKAEPVSECTE-EAERGPAKGDAAEAETAAGALKA 895
 D 449 GDHGLPEYLSPGLPGVDGKGPGLL-CTQCPYIPG-----PGLPGLPLHG 496
 QY 896 KEGSGRATTAKSGAPQDSATCSADVIDEAG-----GKNRLSRPSLTPTGD 950
 D 497 VKGIPGKGAGKSGSPSGNGLPQPPGPAQGPGLKSGETLQPEGGVAV-GD 555
 QY 951 PRANASP-OKPLD-----LKQKORAAIPIQVTKVAPREDAPFKPAPRAPP 1002
 D 556 PGLRGDPGRKGLDGTPTPTGVKGLPGPKGLALSGEKGDGPPDPSPSPGAPG 615
 QY 1003 QNTQESDAPQDQSSPRGSRSPAPADKEAPAEAKQKLGDPDPCWTSGLPFVPPREV 1062
 D 616 -----FOYGPQ-----EPGLQGVGVAP-----GPGGAPGRGE 647
 QY 1063 IKASPAAPDSAPSVAPRHPPLGLHDTARPVLPPTTISNPPLISSAKHPSVLERQ 1122
 D 648 LSVTPVPGPPG-PGPPFHPGFG-----PPGI-----PGLSGKCGDGL 687
 QY 1123 GALSQGSVGLHYPYSEHAKAPVPTMGLPLPMDPKLAPFSGVQEOQLSP--RQAGP 1180
 D 688 -----PDPDGEPIG-----GIGPPGPPGKDGFGFTGKSLGCPGKMEPGL 731
 QY 1181 PESLGPPTAQEASVLRKLTALGVPQGSITKGIPTSTRVPSDAITYRSI-----THGT 1233
 D 732 PGRPLGPA-----KGRPAVAMPDGGTGTGFGEGNSGE-----HGEIGLPLGLPGLT 781
 QY 1234 PADLVYKGTITRILIGBDSPLRDRGSDLPKGVHYEKKGVLYSTEGMSVYQCKED 1293
 D 782 PGNB-----GLDGP-RGDPGQ----- 796
 QY 1294 GRSSGPRHETAPKCTYDMMEGRVGRATISSAIEGIMGRAIPPERHSPHNLKEONHIRG 1353
 D 797 -----PGPPGQPPGRKIEBPRAQG-----LPGLNG-----LGGQGRKG 833
 QY 1354 SI-----TQGPISVYEAQEDYLRRKALKRSESTPPPPSRDLTEAYKTOALGPLKAP 1409
 D 834 KTPGKDPGIRG-----LDRSFPG-----ETGSPG-----IP 861
 QY 1410 AHGELVATVKEAGRSIHETPRELRTPELP-LAPPLKEGI-TQGPPLKDTGASTT 1466
 D 862 GHQEGMPLQGRG-----YFGNPIGLP-PGEDGVIGMVGFP-----GA--- 899

Db 679 DDPAKPVSLP-----PEKTPVLAKKAP-----TKPDSE 710
Qy 1705 QRAMLAGLSPRESSALNTAAGRGITDLSQVHLPLVLPPTGTATMTDLATLPTA 1764
Db 711 AAAPVSGPSSKDKLAKKAPVKPRDPSPKAVPIKP--AKTEVPAAV-----KK 760
Qy 1765 POPSSRRSSSP-----LSPGPTHLTKPTTSSSREDRDRDREREKSILTTT 1820
Db 761 PEPAKSRDPSPKAKKAPNSP--VVPPTPVKNPVK-----795
Qy 1821 VEHAPIRPGTEQSSSGSSGGSSSRASHAHQHSPISPRTDALQORPSVLHN 1880
Db 796 -----WKPMEDDD-----AAEPVNVPEPKTPVLAKTPVKPRDPS-----834
Qy 1881 TGMGIIITAVEPSKPTVLRSTSTSPVRPATPPATHCPLGLDGVYPTLMRVLTPK 1940
Db 835 -PKKAV-----PAKP-----STKTDAF--FVSVKCP-----EFVSKPK 864
Qy 1941 E-APRVARPERPRAD-----TGAFLAKPRASGLSPASPSKSGSPRLVPVSGHATIA 1995
Db 865 EPSKKAEPNSPVVPTPVKNPVKMKRPWEDDDEPTBEVKKPSPEKTPVLAKKEPEK 924
Qy 1996 RTPAKNLAPHHASDPAPASADPHREKTQSKPFSIOLELSLGYHSSYSPEGV-- 2053
Db 925 PKDAPKVA--AKRDPSPKKAPEKEPAKVAAKPRDLSPKKAIPIRANTQEARPTPVKN 981
Qy 2054 -----EPVSVSSPSLTHDKGLPKHLELDKXSHLEGLRKPQGPVYKLGGEA 2100
Db 982 PVKKMKRPWEDDDEPAEVSAP--PEKKT-----VLAK--KAPAKRDPSPKKAAPVA 1032
Qy 2101 A-----HLPLRLPESQP-----SSSPLQTAQVGVGHQVTLAHOISVI 2143
Db 1033 AKDPKPIPEVPTPVKNPVKMKRPWEDDDEPSEVSAPE--EKTTPVLAK--KAP 1085
Qy 2144 TQDYRHHPOQLSAPLPAFLYSPGAS--CPVLRLRP--PSDLYLPP-----PD 2189
Db 1086 TKPRTKPDSEAAAPVSGPTSKDPLSKKAPV-EKPKTTTPKDKLAKSPAKKPEKAP 1144
Qy 2190 HGAPARSGPSHSGGKRSPEPKNT---SVLGGEDGLEPVSPRGMTPEGHSRGAIVPL 2245
Db 1145 PAAPKMKPVWDDPDDEADFTVPAPSKKPDTEPADPLGPK-TXPKLNKKAP--A 1200
Qy 2246 YRDEQTERPSMGSKSPGNTSOPAPAFSKLTESNSAMWKSQKQELNKLNTHNNEPEYN 2305
Db 1201 EKPTKPKPKPV--SKPPKPTPEP-----KPAAPKMKRPWEDD-- 1240
Qy 2306 ISOPGTEIFNMPA 2318
Db 1241 --BEPAD-FTMPA 1250

RESULT 28
15116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 151116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A:Reference number: 151116; MUID:95287814; PMID:7770000
A:Accession: 151116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1110 <JAC>
A:Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632549
C:Superfamily: neurofilament triplet H protein

Query Match 3.0%; Score 403; DB 2; Length 1110;
Best Local Similarity 22.7%; Pred. No. 3.3e-07;
Matches 234; Conservative 153; Mismatches 380; Indels 264; Gaps 44;

Qy 166 RLSEKELLQNMNDVDRREITWVEQOISKLKKOQOLEEBA--KPEPEKPVSPPEI-E 220
Db 97 RVNEKQLLDPLN--DRFAGYIE-KVHDLQGNKLEFETISAYRQKQSGAPAGGVSDIYE 153
Qy 221 SKERSLVQIYDENRKA-----EAARILBGLGROVEL-----PLYNPSDTRQ 265
Db 154 QEIKELRDVIDDINGEKTVOIIEQHLEDEIQRLEKTDDEVRLNTEALIN-----A 207
Qy 266 YHENIKINQAMRKUL-----ILYKRNH-----ARKW-----294
Db 208 FRKNVDUTSLVRMEMDKRTQSLDEITFLKKNHEEVEDLALQIQSSTVSVERDPAVPE 267
Qy 295 -----KQFCORYDLMEALEKKVERIENNRPRAKESVRE 331
Db 268 ITALREIRQLBEGSARNIETAEWFKGKSQUTEAAE-----QNNDAIRAKSEITE 321
Qy 332 YTEK--QPEI-----RKQELQRMQSRVQKQSGLSMAASEHVS---- 373
Db 322 HRRKLQRCETELDLAQTESLERQLSEMERHQSVDG---NLQDAACQOLENELRNTKW 377
Qy 374 EIIDGLSE-QENLEKOR---QLAVIPMLYDADQQRIFKPTNMGLMADPKVYKDRQVM 429
Db 378 EMARHLREYQDLNVKKAALDIEIAVRKL--DGEIRY--SSGLPTPAKPRKAPSAK 432
Qy 430 NMSEQKETEFRKFMQHPKNFGLIASFLERKTVACGLYYLTCKKNENYKSLVRSYR 489
Db 433 PKAKVEKTVYKSK-----PRKYSESIISNQDITDLDLQGEVMEKAA 478
Qy 490 RKSQOQOQOQOQOQOQOQOQPMRSSQEEDEKEKEKAEKEKEKEVENDEKDLKEK 549
Db 479 PVSAEKDEEEEEEKEEKEEAEAEAEEDDRKKEGAEAEEBA-----EEVEKEE 532
Qy 550 TDDTSGDNDKEKAVNAGKRTANSQGRKGRITRSMANVAN--SEEAITPQOQAEI-- 604
Db 533 AEAEVEEAEEAEFEAAAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 592
Qy 605 -----ASMETNESRMTSEEMETAKGLLEHGRWMSAIAMVSGSKTVSQCNFYNYKR 659
Db 593 VEEREAALAEEREEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 641
Qy 660 QNLDEILQHKLXMKERNARRKKKKA PAASAEBAAPPVVED-----702
Db 642 EAEFEVTSKAKAQEAEEVEEAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 701
Qy 703 -EEMEAAGVSGNHEM-----VEEAALHAGNVPRECGSPATVNNSSDTEISPSHTE 757
Db 702 AEDEEAEEVEEYKEEVTKSDAEAEA--EAEEBAKSEBEAEBAEKDEAE--EAEE 756
Qy 758 AAKDTQNGPAPATLADGPPPPPTPRRTSAPITPTPASEATGAPTPPAP-----812
Db 757 AVEETEAATEAEAKESDDEKPEEV---KESAAFPAP---EAKKAPKAPKAPKAP 809
Qy 813 ---PSGAPPPVVKKEKEE--TAAAPVVEEGEQRPAAEELAVDTGAEEBVKSECT 867
Db 810 AKVESPTSEPEDEKAEVKEKGAFAAPK---PKAPPAKAKKAEKVEKEEEDSESP-T 864
Qy 868 EAAEAGAKGDAAATAATAGALKAKKGGSGRATTAASGAPQSDSSATGSADEVD 927
Db 865 EEEBKPPAAKAPK-APAKPPAPKAAE-----KPEPKPAQAPAPAAEED 914
Qy 928 EAEGGDKNRLSPSLITPTGDPANASPOKPLDKLQKORAAAPPIQTVKHEPRE 987
Db 915 EKEDF-----EEEEVEEYKPPDAKPVASKPA-----PAKE 946
Qy 988 DAAPTGA--PPAD---PPONTQPESDA--POQPGSP-----RGKSRSPAP-PAD 1031
Db 947 EEDPEKAKQPKKPKRPAKPEPEEDKAPBAKSHSVBERKPKIKETIAPAKAAPAKAD 1006
Qy 1032 KEAPAAEAOKL 1042
Db 1007 KEPPAAEPKTI 1017

```

Db      1569 GAGMLHITENGPTEV-DY-----SPSDI 1590
Qy      1732 IDLSQVHLPLVPTGTGTPATMDRLAYLPTAPQPPSSSRSSSPLSPGPTHLTKTPTT 1791
Db      1591 QDS-----LSHKLPTEPEPTYQNDLSELISVSQVSPSSAHPSQIASPLQDRTL 1646
Qy      1792 SSSRERDRD-----RERDRERE-----KSILTSTTVEHAPIMRPG-TEOSSGSSGS 1840
Db      1647 SDVPPDMSLYLASLASEKQSLSEKLSPEKSDISPLTPRESSPTTSPGFSDSISGAKES 1706
Qy      1841 SGGGGSSSRPASHASHAHOHSPISPTQDALQORPSVLHNTGMKIIIT-AVED----- 1892
Db      1707 TAAVQTSSPPIDAAAEPYGFSSMLFTWQ-----HHLASRLDITTSVEKNDGKT 1760
Qy      1893 -----SKPTVLKSTSTSPVPAATFPPTATCPGLGTLIDGYPTLMEBVLPLKAPR 1944
Db      1761 PGDFNAYQKP-----ESTTESPDEEDYDSHEKTIQAHVGGYY-----YEK 1804
Qy      1945 VAPERPRADTGHAF-----LAKPARSGL-----BPASSPKG-----SEPRILVP 1987
Db      1805 TERTIAPCOSGSIYETIEKTKTPEDGYSCEITEKTRTPEEGYSYEISEKTRTPE 1864
Qy      1988 VSGHATARTPAKNLAH-----HASDPAPAPASADPHREKTOQKP---F 2031
Db      1865 VSGVTYKTERSRRLDIDISNGYDTEDEGHTLGD-----CSYETTEKITSFPESESY 1919
Qy      2032 SIQELERISLGYHSSSYSPGCVFVSVSFS-----LTHDKGLPKH----- 2073
Db      1920 SYETTTKTRTSPDTSAYCYEMEKITKTPQASTYETSRYCPEKSKSEARQVDLC 1979
Qy      2074 -LEELDKSHLEGELRPK--QGPVK-IGGEAHLPHLRPLESOPSSPLQTAQVKGH 2129
Db      1980 LVSSCFERKHELTSESPFINPFLPMFAGEPTESERPLTQSGCAPPP---SGCKQGR 2036
Qy      2130 QRVVTLAQHISEVITQDTRHHPOQLSAPLPAPLYSPFGASCVP-LDLRRP---PSDLY 2184
Db      2037 QCCEPTPTSVSE-----SAPSGTSDVPERTEECPSITADANLSDSESETIPDKT 2088
Qy      2185 L-----PPPDHGAHPANGSPHSEGGKSPENKTSVLGGEDGIEPVSPREGMT----- 2232
Db      2089 VTYKMDPP--APMO-----DRSPFRHDPV-----SAWDEBALAIEQNLG 2128
Qy      2233 -----EPGHSRAVYPLLYRDEQTEPSPMSKSPGNTSOPAPFSLTES 2278
Db      2129 KALKKDLKERAKTYPGTIKTSSSPVKKGDK-----SKPSAASPKGALKSSDK 2179
Qy      2279 NSAMVSKQOE-INNKLT 2296
Db      2180 VSRVASPKKESVEKAMKT 2198

```

RESULT 27

```

T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U6675; NID:g116613; PID:g116621; PIDN:AA52641.1; GSPDB:GN00
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP.F35A5.1
A:Map position: X
A:Introns: 1272/2

```

```

Query Match      3.1%; Score 404; DB 2; Length 1274;
Best Local Similarity 20.1%; Pred. No. 3,6e-07;
Matches 353; Conservative 167; Mismatches 601; Indels 632; Gaps 87;

Qy      682 KKKKAPAAASEAAPPVVEDEEMASGVSGNEEMVEAEALHASGNEVPRGECGPAT 741
Db      14 KKKMP-----WESVDEEEMEV-----ETPA-----PSKLEKKRSL 47
Qy      742 VNNSDTEISPSRHTTEAKTQNGKPRPATLIGADGPPPPPTPRRTSRAPTE-PPAS 800
Db      48 KKKDAPKVPSP-----DA--PSVPINPVKKWAPWEDDEPME 86
Qy      801 EATGAPTP-----PAPPSAPPPVVPKEKEEETAAPVREGGECQAPPA- 848
Db      87 EAPAPVPKAVKDVDPSPKVPKAPKPRDASPKKIAAKKEPTLAVPTPKVPKPKAP 146
Qy      849 -BEIADVTKAEBPVXSECTEAEABGPAKQDAEAEATAGALKAKEKGGSGR----- 902
Db      147 WEDEVDV-----EDVKDAPTPAKKTPVLKKKPPAAAKPRD--PSPKAAPSKENDPIV 200
Qy      903 -----ATTASSGAPQDSASATCSADEVDEAGGDKN--RLISPRSL-LTPGDPAN 954
Db      201 PPTPIKNPAKWKWPWEDDEVPT--BEIKEBPATKRVAPLAKKEPSTVKPVSDE--- 254
Qy      955 ASPQKPLDLKQLKORAAAIPIQV---TKVHEPPREDAAPTKAPAPAPPPQNLQPSDA 1011
Db      255 -SPKKVPVK---KEPEVPTPIKNPTKKKKPPWEDTEVEYKPEP-----VPEKKA 303
Qy      1012 P--QQPGSSPRGSRSPADPADKFAFAEAQKLPGRPCWTSGLPPEVPREVIKASPHA 1069
Db      304 PVLKKDPAFAAKARDPSS-----KAAPKK 329
Qy      1070 PDSASVYAPRGRPLRLGLHDTARPVLPRLPRTISNP-----ISSAKHPSVL 1118
Db      330 VEPSS-----PVVP-PTPKNPKVKKYKPPWEDDEBAEYKKSAP 369
Qy      1119 ERQIGAISQMSYQLHVPYSEHAKAPVPTMGLPLPMDPKLAPFSGVQEOQLSPRQA 1178
Db      370 EKTTPVLAK-----KEPEPSTTSSDP-SPKKAAP-AVPRROSSPK-KA 411
Qy      1179 GPPEISGVPTAQSASVLRGTALGVSVPGSIYTKGI PSTRVPSDAIYVRSITHTGPADVL 1238
Db      412 TPLQA--DPAQZ-----VPTPVKNPVKKYKPPWME----- 440
Qy      1239 YKGTIRIIEDEPSRLDRGRSDSLPKGHYIYGGKGHVLSYSGMSVTCSEKDGSSS 1298
Db      441 -----VDDEDP--VEEVKQEPAP-----AKTPVLK-----RKEPAKOT 473
Qy      1299 GPPHETAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPRHSPHILKEQHTRGSITQG 1358
Db      474 AKPATSKTPE-----TPKKDPPVPRDSS----- 497
Qy      1359 IPRSYEAQEDYLRRKALKREGTPPPP-----PSRDL---TEAYKTOALGPLELKP 1409
Db      498 -PKK-VAAKPDSAQAPATPVKNPVKKRPPWEDDETPADVDVSKPTAKKTPSLAKKDPAP 555
Qy      1410 AHGELVATVKAAGRSIHETREELRHTPRLPLAPRLKESITQGTPLKTDGTASTGSK 1469
Db      556 AKESLK-----PKADTK--APAKRPDPSPKV-----APTAPEK 587
Qy      1470 KHVRSLIGSPGRTPPVHPLDVMAADALERAICYESLSKSRGTASSSGSIRAGAPVI 1529
Db      588 KTEVLA-----KKEPAGP---ADSKTYS----- 607
Qy      1530 VPELGRQSGPLTYEDHGAFFAGHLPKRGSPVTKREPTPRLQESLSSSKASQDRKLTSTP 1589
Db      608 -PEKSKPRD-----PSPKAAVPAKVPKTEVAPAAVKK-----P 640
Qy      1590 REIAGSPHSTVPRHHHPISPYEHLKRGVSGVDLYKSHIPLADPISIRGI-----PLD 1644
Db      641 EPISK--PKDTAPK-KAEPNSPV-----VP-----PTPVKNPVKKKPPWME 678
Qy      1645 AAAAYVLRHLAENPTYPLYPYLIRGVYDPTALENRQTIINDYITSQOMHNTATAMA 1704

```


A:Accession: 155254
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:g180387; PIDN:AA51993.1; PID:g180388
R:Bonsstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:g180876; PIDN:AA52052.1; PID:g553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter
A:Reference number: 155237; MUID:85130970; PMID:2857713
A:Accession: 155237
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <CH2>
A:Cross-references: GB:M0627; NID:g180383; PIDN:AA51992.1; PID:g553226
R:Wirtz, M.K.; Keene, D.R.; Hort, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A:Reference number: A35233; MUID:90202908; PMID:2318855
A:Accession: A35233
A:Molecule type: Protein
A:Residues: 33-52 <MIR>
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <MBI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4639-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A:Reference number: A30567; MUID:71038625; PMID:5529814
A:Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 162-196, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A:Experimental source: skin
A:Note: evidence for 170-allysine
R:Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F
Eur. J. Biochem. 192, 153-159, 1990
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A:Reference number: S11372; MUID:90382436; PMID:2169412
A:Accession: S11372
A:Molecule type: Protein
A:Residues: 175-187, 274-287, 'P', 289 <BAE>
A:Note: sequence of collagen alpha 1(S) I isolated from bone after pepsin digestion
R:Deak, S.B.; Scholtz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Mnzli, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A:Reference number: 155342; MUID:92042092; PMID:1718984
A:Accession: 155342
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 258-266, 1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:9239007; PIDN:AA820350.1; PID:9239008
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A:Reference number: A92069; MUID:71001508; PMID:4319110
A:Accession: A92069
A:Molecule type: protein

A:Residues: 263-268 <MOR>
A:Experimental source: skin
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A:Title: Segmental amplification of the entire helical and telopeptide regions of the CD
A:Reference number: S15989; MUID:90326017; PMID:2374517
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Precorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A:Reference number: 152905; MUID:93339042; PMID:8339541
A:Accession: 152905
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 342-352, 'C', 354-359 <W12>
A:Cross-references: GB:S64717; NID:g408195; PIDN:AA827677.1; PID:g408196
A:Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A:Reference number: A50476; MUID:84080385; PMID:6689127
A:Accession: A50476
A:Molecule type: mRNA
A:Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A:Cross-references: GB:K01228; NID:g180391; PIDN:AA51995.1; PID:g180392
A:Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiolo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A:Title: Multicase deletion in an osteogenesis imperfecta variant with increased type II
A:Reference number: A22161; MUID:85104934; PMID:2981843
A:Accession: A22161
A:Molecule type: DNA
A:Residues: 472-594, 'R', 596-607 <CH3>
A:Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AA51847.1; PID
R:Walls, G.A.; Starman, B.J.; Zimm, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A:Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A:Reference number: A35336; MUID:90252792; PMID:2339700
A:Accession: A35336
A:Molecule type: mRNA
A:Residues: 710-720, 'E', 722-737, 'E', 739-745 <MAL>
A:Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Fortino, A.; Zolozzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A:Reference number: 154365; MUID:95187161; PMID:7881420
A:Accession: 154365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 746-766, 'S', 768-781 <FOR>
A:Cross-references: GB:I47667; NID:g1009093; PIDN:AA859576.1; PID:g1009094
R:Chesler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A:Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A:Reference number: A47426; MUID:93352646; PMID:8349697
A:Accession: A47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A:Cross-references: GB:S64596; NID:g407589; PIDN:AA827856.1; PID:g407590
A:Note: sequence extracted from NCB1 Backbone (NCBIN:136444, NCBIP:136445)
A:Accession: B47426
A:Molecule type: mRNA
A:Residues: 1179-1464 <CH4>
A:Experimental source: normal dermal fibroblast culture
A:Accession: C47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1464 <CH5>

A:Molecule type: DNA
 A:Residues: 1-1226 <JIA>
 A:Cross-references: EMBL:X59075; NID:94805; PIDN:CAA41799.1; PID:94807
 R:Herbert, C.J.; Jia, Y.; Slonimski, P.P.
 Submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19445
 A:Accession: S19445
 A:Molecule type: DNA
 A:Residues: 1-1226 <HER>
 A:Cross-references: EMBL:X59720; NID:91907116; PIDN:CAA42300.1; PID:e264374; PID:9190717
 R:McKee, B.L.; Roberts, A.B.; Sagliocco, F.A.; Brown, A.J.P.
 Yeast 7, 761-772, 1991
 A:Title: The complete sequence of a 7.5 kb region of chromosome III from *Saccharomyces cerevisiae*
 A:Reference number: S40917; MUID:92133166; PMID:1776366
 A:Accession: S40918
 A:Molecule type: DNA
 A:Residues: 1-304, 'A', 306-374, 'S', 376-403, 'Q', 405-434, 'V', 436-441, 'N', 443-481, 'E', 483-514
 A:Cross-references: EMBL:S78624; NID:9244237; PIDN:AAB21259.1; PID:9244239
 C:Genetic
 A:Cross-references: SGD:S0000629
 A:Map position: 3R
 Query Match 3.1%; Score 412.5; DB 2; Length 1226;
 Best Local Similarity 19.7%; Pred. No. 1.7e-07;
 Matches 236; Conservative 178; Mismatches 438; Indels 343; Gaps 47;
 A:Molecule type: DNA
 A:Residues: 1-1226 <JIA>
 A:Cross-references: EMBL:X59075; NID:94805; PIDN:CAA41799.1; PID:94807
 R:Herbert, C.J.; Jia, Y.; Slonimski, P.P.
 Submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19445
 A:Accession: S19445
 A:Molecule type: DNA
 A:Residues: 1-1226 <HER>
 A:Cross-references: EMBL:X59720; NID:91907116; PIDN:CAA42300.1; PID:e264374; PID:9190717
 R:McKee, B.L.; Roberts, A.B.; Sagliocco, F.A.; Brown, A.J.P.
 Yeast 7, 761-772, 1991
 A:Title: The complete sequence of a 7.5 kb region of chromosome III from *Saccharomyces cerevisiae*
 A:Reference number: S40917; MUID:92133166; PMID:1776366
 A:Accession: S40918
 A:Molecule type: DNA
 A:Residues: 1-304, 'A', 306-374, 'S', 376-403, 'Q', 405-434, 'V', 436-441, 'N', 443-481, 'E', 483-514
 A:Cross-references: EMBL:S78624; NID:9244237; PIDN:AAB21259.1; PID:9244239
 C:Genetic
 A:Cross-references: SGD:S0000629
 A:Map position: 3R
 Query Match 3.1%; Score 412.5; DB 2; Length 1226;
 Best Local Similarity 19.7%; Pred. No. 1.7e-07;
 Matches 236; Conservative 178; Mismatches 438; Indels 343; Gaps 47;

Db 766 DKEKSENNAEVAVQVTVVGVSEVKGDPDLPCTPEKVENMIEKREGEFAGELNENRVNDLK 825
 Qy 562 EAVASGKRTANSQGRKGRITSMANSESAITPQOSAEI-----ASME 608
 Db 826 RAHDEIGEESNKSVALETNNEVQIMAPKGGVNGVYDEETKEHDFSLNALQKKHKSAP 885
 Qy 609 LNSSRWTEBEMETAKKGLIEHGRNMSATARWGSKTVSGCKRPFYNYKKRQULDEILQ 668
 Db 886 EHKTSYWSVRESQFPPELLKEFGSOWSLSEKLGTSITWVNY----- 930
 Qy 669 HKLMKERNARKKKAPAAASEEAPFVVEDEEASGAGNEBEVEAEAL----- 724
 Db 931 -----GRVAAANGKVL-----LVDETLLKRGTS-----SESQSQIILQPE 968
 Qy 725 ---HAGNEVPR-----GECSPATVNNSSDTESI-----PSPHTEAADTGONGP 767
 Db 969 RPNINAVSNIPPOQRALGFVQGPHTGHNTSISIDSGSIRPGDPFH-----BDTESKIS 1024
 Qy 768 KRPATLGADPPGPP 822
 Db 1025 APITTL-----PPP-----RLPSIQPPRSB-----MAEPTVTLNR 1056
 Qy 823 PKEKEEETAAPVEGESEQKPPAAEELAVDTGKAEPYKSECTEAEAGPKADAEA 882
 Db 1057 PLDHIDTLAABAASVTNNQNP---SNERNAIDIGRKSTTISNLNNSDRMKSSFGASR 1113
 Qy 883 AEATAGALKAEKKEGGSGRATTAKSSGAPQDSSTATCSADVDEAGDKRLLSPR 942
 Db 1114 HEAQLF-----DPPSMNIVVQEI-----KENITTPRS 1141
 Qy 943 SLTTPGDPRAANSPOKPLDKOLKORAAIPPIQVTKNHEPPREDAAPTAPR 997
 Db 1142 SSISALLNP-VNNGGSPDGRPLPQHAIS---QGPFPPLAPRTSPISRAP 1193
 RESULT 25
 CGH1S
 collagen alpha 1(I) chain precursor - human
 N:Alternate names: procollagen alpha 1(I) chain
 C:Species: Homo sapiens (hmn)
 C:Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 31-Dec-2000
 C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S115263; A29439; I53466; A02852; I37247
 R:Id Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
 Gene 67, 105-115, 1988
 A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five exons of the human alpha 1(I) collagen gene
 A:Reference number: I60114; MUID:88329734; PMID:2843432
 A:Accession: I60114
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-369, 'L', 371-589 <DAL>
 A:Cross-references: GB:M20789; NID:9179593; PIDN:AAB59373.1; PID:9179594
 R:Tomp, G.; Kuivanen, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock Biochem. J. 253, 919-922, 1988
 A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human type I collagen
 A:Reference number: S01143; MUID:89025644; PMID:3178743
 A:Accession: S01143
 A:Molecule type: mRNA
 A:Residues: 1-472 <TRO>
 A:Cross-references: EMBL:X07884; NID:930015; PIDN:CAA30731.1; PID:930016; GB:M36546; NID:930016
 A:Note: Submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
 R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.; Nature 310, 337-340, 1984
 A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of the first intron
 A:Reference number: A93335; MUID:84270697; PMID:6462220
 A:Accession: A93335
 A:Molecule type: DNA
 A:Residues: 1-58, 'Q', 60-181 <CHU>
 A:Cross-references: EMBL:X00820; NID:935657; PIDN:CAA25394.1; PID:935658
 R:Rosow, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.; J. Biol. Chem. 262, 15151-15157, 1987
 A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enhance transcription
 A:Reference number: I55254; MUID:88033098; PMID:2822714

Db 1828 FRSSMLPTQMHNL-----ALNRDLTSSVSEKSGCKTP-----GDFNVAAYKRP 1871
Qy 1933 MEVLLPREA-----PRVAPRRPADTGHAH-----LAKPPAR 1966
Db 1872 ENAAGSPDEBYDYDESOKRTIRTHDVVAYYEKTERIKSPDCSDSYSEIETKTKPED 1931
Qy 1967 SGL-----EPASSPSK-----SEPRPLVPVSGHATARTAPKXNLAAPH----- 2005
Db 1932 GGYTCETKTKTRTPEEGGYEISEKTRTPEVSGVYEKTERSRLLDDISNGYDDE 1991
Qy 2006 ---HASPDPAPAPASDPHREKTSQKPSIQELETSLGYH-----GSSYSPEG 2052
Db 1992 DCGHTLGD-----CSYSEIETKITSFP-----ESSESYETSITKTRSPDTSAVCYET 2040
Qy 2053 VEPVSPVSSPS-----LTHDKGLPKH-----LEELDKSLHEGLARK--QP 2091
Db 2041 MEKITKTPQASTVSYETSDRCYTTEKSPSEARQPDVLCVSSCEFKHAKTELSPSFLNP 2100
Qy 2092 GPVK-LGGEAAHPLHLRLPESQPSPLQI--APGVKHQHVVTTLAQHISEVITQDYT 2148
Db 2101 NPLEMPAGE-----EPTESB--KPLTSGGAPPSGKQ--QGRQCDETPPISVS 2147
Qy 2149 RHHPQOLAPLPAPLYSPGASCPLY-----DLRPPSDLYL-----PPDHGA 2192
Db 2148 ESASQDSDVDPET-----ECPSTADANIDSESEITPTDKTVYKHMDDPP--A 2199
Qy 2193 PARSPHSEGGKRSPEPKTSVLGGEDGI-----EPVSPREGMTEPGHSRAV 2241
Db 2200 PMQ-----DRSPSPRHPRDVSMDPDALAVQNLGKAVKDKLEKTKTKPKTKSS 2251
Qy 2242 YPLLYRDEQOTPEPSMGSKSPGNTSQPAPFSKLTESNAWKSCKQKINKLNHNNE 2301
Db 2252 SPVKKGDKK-----SKPLAASPFGALKESSDKVRVASPKKASVEKA--TKTTT 2301
Qy 2302 PEYNISQGTETFMNPAITGTGLMTYRSQAOVEHASTMGLAIIIRKALMGKYDQMEESP 2361
Db 2302 PEV-----KATRG-----EKKD 2313
Qy 2362 PLSANAFPLNLSASLPAMPITTAADGRSDHTLISPPGGGKAKVSGRPSRKAKSPAPGL 2421
Db 2314 KETKNA--ANASASKSAK--TAT-----TGPYTKTAKSSTVP-----PGL 2350

RESULT 22
T15348
hypochemical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
A:Accession: T15348
R:Galtung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1; CESP:B0350
C:Genetics:
A:Gene: CESP:B0350.1
A:Introns: 48/1; 503/3; 5116/3

Query Match 3.2%; Score 419; DB 2; Length 5170;
Best Local Similarity 17.5%; Pred. No. 4.8e-07;
Matches 516; Conservative 371; Mismatches 1020; Indels 1040; Gaps 129;

Qy 16 EPRYP-----HSLSPV-----QIARHTTVGLLEYGHSRDYASHL--SPG 56
Db 1345 DQFSPAPSSHVSGIHASESPVAQQQEIPTQRFPHESPAAY--FHEEYESHVLTEQA 1403
Qy 57 SIIGQRRRP-----SILS 70

Db 1404 PLTTEOQHOPESGDSEGLGSKVLGFAKKAQVAGVVAAPVALAAGAKAAYDALKK 1463
Qy 71 EFPQGNRSQELHTRPESHSLPELGKSEMEFIESKRPLELDPDLR----- 119
Db 1464 DDEDEDERESLRLQERSLDSPH--ASESQIEEHEHRRFESVPSKKNHVTETTTT 1521
Qy 120 -----PSPLLA-----TGQPAGE-----DLTKRSLTGKLEPVSPPHPTDP 158
Db 1522 VTRFVDEHEPLVQELIEGKGPADSEKSLPHVVETTTTTTVTRFQKNDSESPVSEK 1581
Qy 159 ELELVPRRLSKBELIQNMNDVDRITWVEQOISKLKKQOOLEBAKPPPEKVPSPRP 218
Db 1582 EOE--PTVSRFVETAKGDEPEHNYETTTTTVTRKVIDDSQEMGDDDEP--KQESQV 1637
Qy 219 IESKHSILVQIIVDEN-----RKAAEAARI-----LEGIG 249
Db 1638 VETTTTTSREYDNDDETSEAGDSHITTTTTVTRFPHGQPEPETETDEVELLP 1697
Qy 250 PQVE-----LPLYNPSD-----TROYHNTKINQAMRKULY 283
Db 1698 PKIEEDNVSEYSSSSTSVSEVRPDEPHIETTTTTVTRFHNPEPETYDDCKDAPI 1757
Qy 284 FKRNHAKQKQRCQRYDQMLEKKEVERIENPRRAKESKREYKQPEIR-- 341
Db 1758 SFSQEH-QDDDSQASHQHDR--ESPVESEKSVKHTTTTTVTRQLYDEASBIRGE 1814
Qy 342 -----KRELOERNQSVRGORS 359
Db 1815 SPVATEEHEVSTKSDSESHVPSVETTTTTVTRFYPDDDELQREDHTQSEERS 1874
Qy 360 GLSMAARSEHEVEEII-----DGLSEONLEK-QMRQLAVIP-----P 397
Db 1875 --SIPTETEHEDHLIKETTTTTVTRFEDPENVKLODSQFSLSPSHVESEIYVP 1932
Qy 398 MLVADQRI-----KFINNGLMADPMKYQKQVQNMMSQEKKE 438
Db 1933 ESPVAKQOEIPTQEFHEDSPAAQVFHEDEYHQVPTQOAPLTLTEOQH--PESGESDGE 1991
Qy 439 TEREFQMRKNGPLGLASFLEKTVABCVL-----YYLTGKN-----ENKSLVR--RS 486
Db 1992 GFGSKVLGFAKKAQVAGVVAAPVALAAGAKAAYDALKKEDDEDEEESLIREERS 2051
Qy 487 YRRRGSKQOQOQOQOQOQO-----OQOQOQPMRPSQEKKE-- 524
Db 2052 FDSPHASQSOILEKHRRFESVPSSEKHNDQSALPQESVSQPLEKSRFFNDESERGV 2111
Qy 525 -----KEK-----EAEKEEKEP--VENDKEDLKEKTDITSGED-- 557
Db 2112 KSDHYTEDDQSLSPKESGEAFSQFTSEKQDRSDSPHISQKD--ISQFONESPDEVK 2170
Qy 558 -----NDEKEAVASGKRTANSQGRKRGIRTSMANEANSSEAITPQOSALASWELNES 612
Db 2171 SEQPHDEKPLEROGSYSGISPKSPGSGITGL--DEKALSGVDEPRDREPNFAS 2226
Qy 613 SRWTEEMETAKGLLEHGR-----NMSAIAVMGSKTVSOCKNFFYKGRQNLDE 664
Db 2227 HEKTB--ATSDENLFBSDKTAAPASPVPSSEDSNRVIEITTTTTVTRHFPEDDHSYV 2283
Qy 665 ILQGHKLC--MEKRNARRKKKAPAAA-----SEPAAPPVYED--EEMESGVSGN 713
Db 2284 ESQEVSSSGSPVPEKSVDRVIEITTTTTVTRHFEADELPIITVESHHDQOASFPV-S 2342
Qy 714 EEMVEEHAL-----HAGNEVPRGEGSPATVNNSS 746
Db 2343 EEDVHQIQOTTTTTTVTKHFVPDDEIDESHMNESDKYAGSPVPSSEDSRVYETTTT 2402
Qy 747 DT-----ESIPSPTEAAKDQGNPKRPPA-----TL 773
Db 2403 TTVTRHFPEDDHSFPVQTOYASSESPVSEKSVREVITTTTTVTRHFEDEHIL 2462
Qy 774 GADGPPG-----PPTPRRTSRADIEPT-----PASEATGAPT-----PP 810
Db 2463 GQGGESDQIPSESITSEMDRETSSSPVQNRDEEFVLPAIAPYKQCTEYGRVDSHAP 2522

F:1861-2064/Region: 17-residue repeats
 F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: pH
 F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (thr) (cd
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 3.2%; Score 424.5; DB 1; Length 2464;

Best local similarity 19.3%; Pred. No. 1.4e-07;

Matches 521; Conservative 314; Mismatches 996; Indels 869; Gaps 114;

```

QY 68 LLSFQGNERSQELHLPESHSLVPLG-----KSEMETESKRPLELLPDL 118
DB 174 LLSFTHPANKASLTL-----FCPEGDWKNNSLDRHNLQDFINIKLNASLTP----- 221
QY 119 RPSFLATGCPAGSEDLTKORSLGKLEVPSPSPHDPLELVP-----RLSKR-- 170
DB 222 -----EMEGSLSEFTEYLS-----ESVSPSPFDI-----LEPPTSGFLKLSKPC 262
QY 171 -----ELIQMDRVREITVEQOI----- 190
DB 263 YIFPGKGSALFVNGFNMLINGSEKSCFWKLIRLDRVD-----SILLTHIGDDMLPG 319
QY 191 --SKLKKKQOOLEEBAKPEPEKVPSPPIESKHSRLVQIITYENRKGAFAHRIIEGL 248
DB 320 INSMLOKRIAELEBE-----RSQGSTNSDMKXLI 350
QY 249 GPQVELLYNQSPDTRQYHENIKINQAMRKLLIFKRRNNAKQWQKQFCQRYDQLEMA 308
DB 351 SPDLGVFLNVPENLKDBEPNFKMRKSTIEACFTL-----QYLNKSMKEPPLFRS 401
QY 309 LEKVE-----RIENPPRRRAKESKREYVEKQPEIRKOR--ELQERQSGRVGQ 356
DB 402 VGNTEFVILFQKNVGKLEMYVLNPKSSKEMQYFMQWGTGKDAKELLPRNGQEVDI 461
QY 357 RGSGL-SMSAARSHEVS-----EITDGLSEQNTLEKQKQALAV----- 394
DB 462 PISLTYSVSLIWMHPANPAEKIRVLPPGNSTOYNILEGLEKHLDFLQPLATQKDL 521
QY 395 -----IIPMLYDAAQOQIKFINNGMLADPMKVVKQDQVNMMSQOEKE--TFRKPMQHP 448
DB 522 TGOVPTPKVQVTKQR--ADRSRLKPTKPVASVKRESKELEPTEVTKTSQVETPR 578
QY 449 KNFGLIASFLERKTVAECVLVYVLTKNENYSKLVRSYRRRGKSOQOQOQ--QOQOQOQ 506
DB 579 K-----VESKRV-----LVKKDKPVKTESKSPVTEKEVSKNEQSPVAAVAEK 623
QY 507 QOQOQPMRPSQOEBQDEK--KEKAKEKEKEPEVENDEKDLIEKTDTSGEDNDEKAIVA 565
DB 624 QATESKPKVTQDKVYKKEIKTKLEKKEKEPK--KEVVKEDKTPLKDKBEKREKEVK 679
QY 566 SKGKKTANSQGR-----KGRITSMANENSEEAIIPQOQSAELASMEIN 610
DB 680 KEIKKEIKERKRLKEVKEKTPDKAKKEVKKEKKEKKEK--BEKKEIKIKISQDIK 738
QY 611 ESSRWTEEMETAKGLLEHGRNMSAIAWVGSKTVSQCKNFYFNKKQKQMLDILQOHK 670
DB 739 KST-----PQSTTKP-----SALKPKYAKKEESTK-----EPILAAGK 772
QY 671 LKMEKERNARRKKK---APAASEEAAPPVVEDEMEASGVSGNEEWEAEALHA 726
DB 773 LKDGKGVKVIKEEGKTEEAATAVGTAAVTAALAAAGIASG-----PVKELEA--- 822
QY 727 SGNVPPGEGCGPATVNNSSDTEIIPSPHTAADQNGPKPATLGDADPPGCPTRP 786
DB 823 -----EBSLWSSPDDL--TKDFEELKAEIIVAKDI--KPOULELD----- 860
QY 787 RRTSRADIEPTPASEATGAPTRPPAPSPAPPVVPEKEKEEETAAPVE-----EGE 841
DB 861 -----EELKETQPEAY-----VIQETEVSGKSAESPBGITTTTGE 899
QY 842 EQKPAABEALAVDTGKAEEPVYSKTEBEAGPAKGDAAEALATAGALAEKKEG--- 898
DB 900 GECEQTEBEL-----EPVEKQGVDDIEKFEDGAGFEESSEFGYEEKALTEEAEEP 951

```

```

QY 899 ---SGRATTAKSAGAFQDSDSATCSADVEAEGDKRNRLSPRESLTLPTGDPANA 955
DB 952 EEOEDNDASSAGSHSTEDDESKAADV-----HLKERSV--SCDDAAEE 999
QY 956 SPQKPLDKLQKQRAAIIPIQVTKHEPPREDAAPTKPPAPPPQNTQPSD----- 1010
DB 1000 DMDVLEKGAEOG-----EEGSEEDDAEDAREGEYEPDKTEADYVAAVADKAFA 1052
QY 1011 -----AQQRC--SSPKCKSRPAP-----ADKEAFAEAQKLGQDPC 1048
DB 1053 GVTEEQYGLGTSKQGIQSP--SREPASSIHDETLPGSSSEEAATADSEENEDQPE 1109
QY 1049 WTSGLPPVPPEVITKASPHAPDSASVAPPGHPLPLGLHDTARVPL--RPPTISN 1104
DB 1110 FTASGYTQSTIEI-----SSPFTMDMSSTRDVMSETNNETESP 1152
QY 1105 PPLISSAKHPSVLERQ-----IGAISQMSVOLHVPYSEHAKAVGVPTWGLPLP 1155
DB 1153 SQEFVNITKYESLSYQGEYSKPAVASFNGLSEGSTDATQDKDYNASAS---TISPSS 1208
QY 1156 MDPKULA-----PSSGVQEOQLSRGQAGPPESLGVPTQAQASV 1194
DB 1209 MEEDKFSKALRDAYCSEKEKELKASAEIDIKVDSERLSP--AKSPILSPSP-----SP 1261
QY 1195 LRGTALGVPGSIITKGIPTSRVPSDAITYRGSIITGTPADVLKGTIRIIGEDSPSR 1254
DB 1262 IETKPLQE-----KSNVSLTPNEIKVAGAEARVSP-----GVTAQVVEHCASP 1308
QY 1255 LDRGREDSLPKGHVLYEGKKGHVLVEGMSVTQCSKEDRSSGP---PHETAAPKRT 1310
DB 1309 ECKTLEVAVSPQSV--TGSAGHTPYQ-----SPTDEKSHLPRVESENAQAVP--- 1355
QY 1311 YDMMEGKVGSAISASIEGLMGAIP-----PERNSHHLKEQNHITGISTQGIIPRYVE 1365
DB 1356 -----VSFEFSKADENERASLSPMDEVPQSESEVE--KVLSPILSPILLGSESP-- 1405
QY 1366 AOEYLREAKLKR-----EGTPPPPPPRDITAEAKYTOALGPLKLPAAHEG 1413
DB 1406 --EDFLADSKVLGRBESPEEGKNGKQGFPPDRSPVSDLT---STGLVODKQEKSTG 1459
QY 1414 LVATVKEAGRSIHEIPREELRHTEPLRLAPRLPKESGITQGTPLKYDTGASTGSKKDV 1473
DB 1460 FIPKEDEFPEKKTSDVETWSSGALALDERKL--GQDV---SPTQID--VSQFSGFKEDT 1513
QY 1474 RSLIGSPGRTPPPPHPLDWMADARALERACYEBSLKSPGTASSGSGSIARGAVIYBEL 1533
DB 1514 KMSISEGTVDKSAITPD-----EGVAEDTYSMEGVAASVSTASVATSS--PPEP 1561
QY 1534 GKPRQSPUTYEDHGAAPAGHLPRGSPVTMRBEPTRLQEGSLSSSKASQDRKLTSTPREI- 1592
DB 1562 TTDDVSPSLAIVGSPSTEVDSLSVSVQTPPTPQETEMSPKESCPRMSISPPDFS 1621
QY 1593 ---AKSPHSTVPEHHPHPISEYHLLRGVSGVDLYRSHIPLADPTSIPRGIPLDAAAY 1649
DB 1622 PKTAKS--RTPVQDHRSE-----QSMSTIEFGQESPEHSPAMDFS--- 1659
QY 1650 YLPRHLAPNTPYPLPYLIRGYPTAALBNQTIINDYITISQGMHNHNTATMAQAADM 1709
DB 1660 -----RQSPDHPLT-----GASVLHITENGTEV--DY----- 1665
QY 1710 LRGLSP---RESSALNLYAAGPRGIIIDLSQVPHLVLVPPTPGTPATAMRLAYLTPAQ 1766
DB 1686 ---SPCDIQDSSLSHK-----IPTEBESYIQDNDLSLISVSQ 1721
QY 1767 ---PSSRHS---SSPLSPGPTHLTPTTTS-----SSERDRDRERDRERREK 1812
DB 1722 VEAQSPSTSAHTPQIASPLQEDTLSDVPPREMSILVASIASSEKVCGLGEG---KLSPK 1777
QY 1813 SILSTTTVEHAPTRMRCGTQSSSGSSSGGSGSSSRPASHAHQHSPISTPTQALQ 1872
DB 1778 SDISPLTPRESSPLYSYGFDSDSTSAKETAAAHQAASSP-----PLDAATAAPVYG 1827
QY 1873 QRPVSLHNTQWKGIITAVEPSKPTVLASTSTSPVPRPAATFPATHCPLGLDGVPTL 1932

```


QY	807		PPPPAPBSAPAPPPVPPKKEKEEPTAAAPRVBESEQCPRAABELA-----	852
Db	88	SPPPAPSPD-PDGPAPSGAGAAE-AGTPPSQCPAGTTPPSQGAAGKGDGTAPSGT	145	
QY	853	---VDTGAAEEVNSECTEEAEDEGPAGKDAEAAAEATAE-----GALKAKEKGGSGRAT	904	
Db	146	KSGADGKPAQGVPRATTATATBARABASAAPVPRKATENATAVTAASGAPDAADDAAV	205	
QY	905	TAKSSGAPODSDSSATCSADEVDEABGDXRLSBPSSLTPTGDPANA-----S	956	
Db	206	TAAOSGAPKATVEVPRAAAANAANKA-----AVTAAAAAPKATAKAPAVYS	253	
QY	957	POKPLDKOLKORAAAP-----PIQVTKVNEPREDAAPTCKAPAP	1000	
Db	254	PTTPSSAPAKPLTAASPTASKATBAKAVPRATASIMATKV-----TAEKPA-PSBS	305	
QY	1001	PPONLQPSDABQOGSSPBRGSRAPAPRADXEAPAAEAOXLGCDPCCWTSGLPPVPR	1066	
Db	306	VP---KATDTKAVATAPAK-----AGDVPAAVAVCAEAPAPPP-----PQ	346	
QY	1061	EYKASPHAPDASATYAP---PGHPLPLGLHDTARPVLPRPTISNPPPLSSAKHBPV	1111	
Db	347	QUPKAA-AAAATGTELKPAATAPPPGSP---RANSHTVVTTPNV---PRAAAATVPA	398	
QY	1118	LERQIGALISQGMVOLIHPYSEHAAPVGPVTMGLPLPMDPKLAPFSGVKEQOLSPRQ	1177	
Db	399	---GAVPKAST-----GTPRAAPQOPV-----PKAAPVTPPSQOAVPR---	435	
QY	1178	AGPSSLGVPPTAOEASVLRGTLGSPGSGITKGIAPSTR-VPSDAITYRGSITHTGPAD	1233	
Db	436	---AATPAAAPV-----TPQOPVTKAATTTNATPPOPPIPKAATTTATAPV-	478	
QY	1237	VLYKGTITRILIGEDSPSRLDRGREDSLPRGHVITGKKGHVLSYGGMSVTCSEKDEGS	1296	
Db	479	---TP-----QQPIPA-----GTD	490	
QY	1297	SSGPRHETAAPKRTYDMMEGRVGAALISASIEGLMGRAIPPRHSPHLKEOHIRGSIT	1356	
Db	491	AAPPP---AVPRAPSD-----GRAATP-----	509	
QY	1357	QGISVYEAQGDYLREAKLKRGGTPRRPPSRDLTEAYTQOLGPIKLKPAHEGLVA	1411	
Db	510	-GVPPAATDPQK-----PPTQOS-----	527	
QY	1417	TYKEAGRSIHLEPREELRHTPEPLAPAPRLKESITQGT---LKUDTGASTTGSKKDV	1477	
Db	528	---VPSAVTEPKQPPRAAPPSNATAPVPSPNLK-----	561	
QY	1474	RLSLSPGRTPRPVPRPLVMDADARALERAQESLSKSRGTASSSGGSIARCAPVITPEL	1533	
Db	562	---SPLETTIP--KEVPLMA-----LTPOPVTAQO-----MVQOL	589	
QY	1534	GKPROSLPTYEDHAPFAGHLPRGSPVTWKEEPTPRLBOESSLSSKASODKLTSTPREIA	1593	
Db	590	AATKSPIT-----VKAAPKALMTPPP-PRGLPRALAAAKLGLPSSP--VA	634	
QY	1594	KSPHSTVEHHHPHPIPSYEHLRLGVSGVDLYRSHLPAFDPTSPRGIRPLDAAAAYLPR	1655	
Db	635	SAMAHAKVTP-RPLPASP-----VPMASPASLGP---DAARV-----	667	
QY	1654	HLAARPYRPHLYRPLINGRPTALLENKQTIINUYIISQOCHNHNTATAMAQRADMGL	1711	
Db	668	ALATPAASGAKP-----EAGANGTMAAMG-----	694	
QY	1714	SPRESSLALNYAAGRGIIIDLSQVPHLPVL---VEPTGTPATAMDRLAULYPTAQPPSS	1770	
Db	695	---AANTGMARITGAAGAQTAPMGAANHNVSPMAGCATQW---SPTGAAN	739	
QY	1771	RHSSAPLPGCGTTHLTKPTTSSSERDRDRDRDRDREREKSIITSTTYVHAANIWRPG	1833	
Db	740	TH-MSPIAGGAGTQWS-PWGAANTQMSR-----MGATITQMSPWGAA	780	

QY 1831 IEQSSGSSGSSGGGSSSRASHSHAHQNSPIRQDOLQDRPSVLNHTGMKGIITAV 1890

Db 781 TTQSPSPM-----GAAAGVATATSAAGNTMOVSPMGAAATP---PQTPSV-----GAATTP 825

QY 1891 EPRKPTVLSTSTSSPPVPAPATPPPATHCPLGGLDGVPTLMEPVLLPKKAPAVARPER 1950

Db 826 QPS--PMGAATLMSPMGAATTPQPS--PWGAN-----TTQPP--PMAATTTQPPP 871

QY 1951 PRADT-----GHAFLAKPPARSGLEPASSPSSKSEDRPLVPPVSGHATTAATPAKNLAP 2004

Db 872 MAASIPQSTPMGAATTTQSPPMGAATTTQSPPMGASTPQ--APF-----TYAGSP----- 918

QY 2005 HHASDPDPAPAPASADPHREKTQSKPPSIQELRLSLGTHGSSYSPEGVPEPVSSPSL 2064

Db 919 ---TPPPPIPPSPPTA-----QTSPPQMS-----KSPPPDP-- 945

QY 2065 THDKGLPRHGLEEDKSHLEGELRPKQPGPVYKGGSEAHNLPHLRPLPSQSPSSPLDGTAP 2124

Db 946 -----PKAPSAAAQTSPALHVNAN-----ASP 966

QY 2125 GYKGHQRYVTLAQHISEVITQDYTRHHHPQOLASAPLPAPLYSFPGASCPLDLRRPPSDLY 2184

Db 967 GV-----TANSPAPI----- 976

QY 2185 LPPDHGAPARGSPHSBEGKRSPEPPNKTSLVAGGEDGIEPVSPREG--MTEPGHSKAV 2241

Db 977 -----GVTEASPSADAGARLSFGPTAAT-----DG--PKASPAATADVTEAATDTTAA 1021

QY 2242 YPLLYRDEQIEPSPRMGSKSPGNTSOPPAFESKLTESNSAMVSKQEIINKUANTHNRNE 2301

Db 1022 ATAPAPAEAPAPKARSSSSSSSSSS-----SSSSSSSSSSSSSDSSSSSSSESNPAS 1075

QY 2302 PEVINSQPGTEIFPMNPATITGTMLTYRSQAVQEHASTMGLLEAIRKALMGKYDQWESP 2361

Db 1076 P-----APAVG-----DGGQQMT 1088

QY 2362 PLSANAFENPLNASSLPAPMPTIADGRSDHTLTPSGGGGKAYKSGPSSRRKAKSPAPGL 2421

Db 1089 PGAAQSVPTVTEAAVQEHAAAAAAG-----AEREGRPTRRKRTTRSSSS 1134

QY 2422 ASGDRPPEVSSVHS 2435

Db 1135 SSSSSSSSSSSSSSS 1148

RESULT 21

QMSMRI

microtubule-associated protein MAP1B - mouse

N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 01-Sep-2000

C/Accession: S075549; S44387; A33645

R/Noble, M.; Lewis, S.A.; Cowan, N.J.

R/Cell Biol. 109, 3567-3376, 1989

A/Title: The microtubule binding domain of microtubule-associated protein MAP1B containing

A/Reference number: A33645; MUID:9094539; PMID:2480963

A/Accession: S075549

A/Molecule type: mRNA

A/Residues: 1-2464 <NO>

A/Cross-references: EMBL:X51396; NID:952999; PIDN:CA53761.1; PID:953000

R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabaia, J.C.; Avila, J.

Arch. Biochem. Biophys. 310, 428-432, 1994

A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.

A/Reference number: S44387; MUID:94234720; PMID:8179328

A/Accession: S44387

A/Status: Preliminary

A/Molecule type: Protein

A/Residues: 653-663, 'IC' <SAN>

C/Keywords: microtubule binding; phosphoprotein; tandem repeat

R/589-786/Domain: microtubule binding #status experimental <NTB>

R/589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-693

R-K-E/D-X)

```

Db 434 GPNVNRQVSHTPGAALVNLALRPSPHHTFLGHSBPNIAKTXAISREHLKIQFNS 493
Qy 1107 -----PLISSA-----KRPVLERQ-----IGAIQGMS-VOLHV 1135
Db 494 QAGVFPAIPLHKNGFCEVDVHYSHDKVLSGDLQVKDVEFVEIINGVAGKGAEEYE 553
Qy 1136 P-----YSENAK-----APGVPTMGLPLMPDKPKLAPFSGVQEOLOSP 1174
Db 554 PEETPARAYSEGGEKMSFDESHIDVDRKSTSPEDNEVMSVDDR-----DQSQELS- 607
Qy 1175 RGQAGPESLGVPAAQE--ASVLRGTA-----LGSVPGG 1206
Db 608 -----EPPEDELLPDPAPHMETVEKEAEEDDDEARSQOSVKEPPEPPEPDMSSVPLM 664
Qy 1207 SITKGISTVPSASATYTGSIHTGTBVDLYKGT-----ITRIIG----- 1248
Db 665 DPPKRGPRGPRPKNGIMSKEERLRKKQAMELAKNOPPPQPGSPPKRKGPRKHP 724
Qy 1249 EDSPRRLDRGREDLPGKHVLYEGKKGHVLSYEGGMSVTQSGKEDGRSS--SGPPEHTAAP 1307
Db 725 EDAPDRPEKKKPKYKRRK-----NGEEDDASDAE-----KTIKREKREKTPLELRRE 773
Qy 1308 KRTYDME-----GRVGRALISSASIEGLMGRAIPRRHSBPHLKEQHNRGSIITQGI 1361
Db 774 DYTEBQLOKPNKNGVGLIDEVLSAAPDGLTKOI-----YKRIQKPYFYFNVDTKGMS 829
Qy 1362 S-----YVAQEDYLKREKLL-----KREGTTPPPPPS--RDLTAYTQOLGP 1404
Db 830 SVRRNLIGNDAPKKNBETHLMSRVPGIDIDAGKKRKAAPSPDHASLNFQGHVAPQPM-- 887
Qy 1405 LKLPKAEGL-----VATKAGRSIHETPREELRH-----TPE 1438
Db 888 -----PHRGMVHGHGVQOSYHPTGVQOROSVYTTGQPGASQH--PQHLOTTPPGVPRQ 941
Qy 1439 LPLAPRLKSGSITQGTPLKYDTGASTGSKKHVRSLGSGRTF-----PVV----- 1487
Db 942 QP--PRRAYQAAQTSPPAQ--AQYGRPTPAQMGSTPRATYSSPYVSRPMTVA 995
Qy 1488 -----HPLDVM--ADARALERACYESLSKSRPTASSSGSITARGA----- 1526
Db 996 AOSGATPRMARQSLPVSQSGSQAINGIPIV-----NPPATANTPVAVAGGARPAQA 1048
Qy 1527 -----FVIYBEL-----GKPROSPLTYEDHGAPFAGHLP-----RGSPTVM 1562
Db 1049 TPTANTTAAPRANLVIAPELISWLESFKVYKLEYIQTSKFPQIILAMSVINGLKITT 1108
Qy 1563 REPTP-----RLQ-----EGSLSSRAS- 1580
Db 1109 KSMIPDESELVVLVRFVEERIOGTSKSLDPLDQLTLTPFKATWSTLEAKLDQSQAEC 1168
Qy 1581 -----QDRKLT-----STPRE-----IAKSP-----H 1597
Db 1169 LVLSAIDQVLGLADKITTRGTESMRBFNNAEKYLIPAIRMKVAMORKOVAATPAAPVH 1228
Qy 1598 ST-----VPEHNPRISSP--YEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGILPAAAYY 1650
Db 1229 ATATPGLPLPNH--HTMAPATAPSAQRIATPISATQORANPCGANSAPRTITTAAPR-- 1255
Qy 1651 LPRHLANPTYPPLHYRPGYLDTALENROTI-----INDYTSQOMHNTATAMA 1704
Db 1286 -----TPAPILPRAAP--ISGHDIAPSTVNTNAAMGARVPVAPVPOQGVHSGAIG 1337
Qy 1705 QRADMLRG-----LSPRESSALANVAGRGITDLSQVH--LPV----- 1742
Db 1338 VTPAVSSPPTVTHGVTAAPTTTPR--SNLASVAVG-----SMVQOQTHSGAPAGSGNAST 1391
Qy 1743 -LVPTTGTATAMDLAVLPTAPQPPSSRRSSPLSPGCTHLLTKTTTSSSERDRD 1801
Db 1392 SRAPPSIAAPTA-----PTSVPVSVSTYRPMHSYVPTGPGGILTAFTTASSG----- 1446
Qy 1802 REPRDREREKSIITSTTYEHA-----PIWRPQTEOSSGSSSGSGGSGSSSRPA 1852
Db 1437 ASAGYAPANASTWPAFTSGQAAMTSAVPOGVSPRP--SLITQOMGVPAIATAASTSRPA 1495

```

```

Qy 1853 SHSHAHQHSPISPRTODALQORPSVLNMGKGIITAVE-----PSKPTVLR--STSS 1905
Db 1496 SGVINPASSLABSTHKM--PSAVPTTASGVSSVSSLATPLPSPAPRQSPSNAT 1552
Qy 1906 PVPRATFPPTATCPLOGTLDGVYPTLMEVLLPKENAPVARBERPRADTGHAFAPRA 1965
Db 1553 PTPPATAPRA-----PL-----PAASVASPVYQGP-----SFA--PA 1585
Qy 1966 RSLGPASSPSKSEPR--RPLVPVSGHATITATPAKMLAPHNASPPRAPPASDPHR 2023
Db 1586 SVPTPTPSAASQAOPLSQCPGAPPAISSISGATPAASI-----PSSA----- 1628
Qy 2024 EKTQSKPFSIQEELMSLGVHSSYSPEGEVPSPVSSPGLTHDKGLPKLBELDKHLE 2083
Db 1629 -----PALAPVT----- 1636
Qy 2084 GEIRPKOPGFVKLGEMAHILPHLRPLPE-----SOPSSSPLQTPAGVKGHORVTL 2135
Db 1637 -YVPQOASAAARLP--VTPAPAAHTIAQSVAPVAPRVTOGP-----VQSV 1681
Qy 2136 AQHI--SEVITQYTRHHPOQLSAPLPAPLYSFPGASCPLYDLRRPPSDLYLPPDHGAP 2193
Db 1682 AOHVTSQAST--TAAP--VAQSVPRPVNSPTSAAPV-----A 1717
Qy 2194 ARGSPHSEGGKRSPEPVKTSVLGGGEDGIEPVSPPEGMEPRGS--RSAYVPLLYRDEQ 2251
Db 1718 AVGAQVASAPTYTQPPRHALSSVSQSL--POSVPAAQOALQIOTPSASRPV-----PQ 1771
Qy 2252 TERSRMGSKSPGNTSQPPAFPSKLTBSNSAMVSKKOEINKUNTNHNEPEYNISQPGT 2311
Db 1772 SVPOSVQATPPQAVPR-----STSLTLPT-----AQGP 1801
Qy 2312 EIRMPAITYTGMLTYSQAVOEHASTMGLKAIIRKALMGKTDOMEESPLSANAENPL 2371
Db 1802 V--SPAVSGSGVPA--PSAAQGVAPAPV-----SSTPVPAATVAPA 1838
Qy 2372 NASASLPAAMPITPAADGRSHTLTSPGCGGAKAVSGRPSRKASPAAPGLASGRPPSVS 2431
Db 1839 STVAAAPTPTRVTAAPALAS-AATNP-----APVPSOP--OHQTTGQAP--AQOORPPAQA 1889
Qy 2432 SVHSEGDGCRRTPLTNKRWEDRPSASGTFPPYNPILMRLQAGVMSPPPGGLPAGSGPL 2491
Db 1890 QAPR-----TPTTISAAPRRPPTLAPPPPPPP-----PTEDEPPPPPPPPAPAP 1936
Qy 2492 AGPHAMDEEPKPL 2506
Db 1937 PPP-----PTPLM 1944

RESULT 20
T18535
high molecular mass nuclear antigen - chicken (fragment)
C/Species: Gallus gallus (chicken)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T18535
R/Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A/Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chicken
A/Reference number: Z18955; MUID:9803440; PMID:9365273
A/Accession: T18535
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1151 <SH1>
A/Cross-references: EMBL:DB8440; NID:d1177138; PID:d1025045; PIDN:BA24137.1

Query Match 3.3%; Score 436; DB 2; Length 1151;
Best Local Similarity 19.7%; Pred. No. 2, 4e-08;
Matches 346; Conservative 154; Mismatches 548; Indels 706; Gaps 77;

Qy 752 PSPHTFAKDTGONGPKRP--ATLGADGPPG--PTPPRR--TSRAPIETPSAEGATGAP 806
Db 31 PRKWPVIAELHAAQPPPKWPIGAPPPPTBTPPKPTDGADAPKASAEIT--- 87

```

[illegible][illegible]

Db 2875 VSLIAGRLSGSPSDL-----QNHVAGSGGERSAGDSQPRNPPTAQCIVINADDOQY 2930
QY 2228 PEKTEPGEHRSRVAVPL-----YRDEQTEPBRMSKSPGNTSQPPAFSKITES 2278
Db 2931 EEWLF--HTQQLQOMOLKYLEEQIGVHRKSRKALKAKORTAKKAGREFEADAEK- 2984
QY 2279 NSMNVKSKQKQINKLMTNHNREPEVNIISQGTETFMMPAITGTGLMT--YRSQAVQEHAS 2337
Db 2985 --KLVEEQSKTIQKQLQVRRQCKEH-----TNLAEYRNKQKQKQ-- 3023
QY 2338 TNNGLAEIRKALMGKYDQWESPPLSANAFNPINASASLPAAPIYAADGRSDHTLTSP 2397
Db 3024 -----QQQQQQQQHSAVLAISP--SQSPRLITKLPQGLLRG--HGIQRP 3064
QY 2398 GGGGKAVKSRPSSRKAKSPAPGLASGDRPPSVSSVHSEGDGCRRTPLTNKRWEDRPSA 2457
Db 3065 QG-----PPGQAGG----- 3074
QY 2458 GSTPFPNPLIMPLQAGVMSPPRPLP-----AGSGPLAGP 2494
Db 3075 -----LRLTPGGMALPGPGGPFNTALAAQQQQQHSAGSLAGP 3115

RESULT 18

T03454
ALR protein - human
C/Species: Homo sapiens (man)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 27-Oct-2003
C/Accession: T03454
R/Parad: R.; Zhdanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A/Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A/Reference number: Z14954; MUID:97388474; PMID:9247308
A/Accession: T03454
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-5262 <PRA>
A/Cross-references: EMBL:AF010403; NID:92358284; PIDN:AACS1734.1; PID:92358285
C/Genetic:
A/Gene: ALR
A/Map position: 12
C/Superfamily: acute lymphoblastic leukemia protein, ALR type
C/Keywords: alternative splicing

Query Match 3.68; Score 470.5; DB 2; Length 5262;
Best Local Similarity 20.3%; Pred. No. 7.6e-09;
Matches 558; Conservative 284; Mismatches 891; Indels 1019; Gaps 144;
QY 145 LEVUSPSPHTDELIVPRLSKE-----LIONMDVDEITVNEQISKLK 194
Db 1286 VKPVAVAP-----ELVPMKVKEPBPQYFRFGWMLTETGMALNNLT-SPH 1334
QY 195 KKQOQ-----LEEEAA-KPEPEKPVSPPIESKHSILVQIYDENRKAFAARILEG- 247
Db 1335 KRRQRRRLGLPGAGJGSESPDALGRDDKODLDTDELKKEGVEHNECEIKIEGR 1394
QY 248 LGFOVELPLYNQSPDTQYHENIKINQAMRKL-----ILYKRNHAKQWKQKFC 299
Db 1395 VSPDVE--PGKEETESKK-----RKRKPYRPGIGGFWMQRKSHRTK-KPPAA 1441
QY 300 QR-----YQLMALAEKXRYETIENNRRAKSKVAYEYEQ 336
Db 1442 QAEVLSGDGPDEVITPADLPBEGAVEQSLAEQDEKKQO-----RRGRKSKSLGMPAY 1496
QY 337 FPRIRKORELOERMQR-----VGORSGLSMSAAREHEVSEIIDLSEQENLEKQMRQ 391
Db 1497 LQEFKFKELLDLSRKALFAVGVGRPSGLGTPKAKG-----DGSSEKELPTS--Q 1546
QY 392 LAVIPMLYDADQORIK-----FINNGMLADPMKYKQROVNNMMSBOEKETFRREKFM 445
Db 1547 KGDGDPDIADBSRGLGKADTGPEDGQVKAISVP-----SDPEK----- 1587

QY 446 QHPKNFLIASFLERKTVACVLYYYLTKKNENYKSLVRSYRRRKGQQQQQQQQQQQ 505
Db 1588 PGTPEBEMLSDDLDRISTEE----- 1607
QY 506 QQQQQPVRSSQOEKDEKEKEKE--AEKEE--KPEVNDKEDLLK----- 547
Db 1608 -----LPK--MESKQQLQPKDVLGSEREQHGCPTGLEGSRTPLOQPLQGLN 1659
QY 548 -----EKDDTDSG-----ENDKEAVASGKRTANSQGRKGRITRSMANSEASAIT 597
Db 1660 LPSSPMDSYPCQSPFLDSRERGFFSPSPGPPSPMTGSGTTPSTPT--T 1711
QY 598 PQSASIELASMEINNESS--RWTEEE--METAK-----KGLBHRNMSAIARWVG 642
Db 1712 PTTBEGSDGLSYQRSIQRMKDEBELQGLSTIPVLYANINFPRLKDDYDWS----- 1764
QY 643 SKTVSQCKNFYFNKKRONLDEILQHKLMEXERNARRKKKAPAAASEEAAFPVVED 702
Db 1765 -----SRCKQIMKLMRKVPADKAPYLQKAK--DNRAHRINKVQKQAESQ-----IN 1810
QY 703 EEMBASGVSGNEEMVBEAALHAASGNEVPRGSCGATYNNSDTESISPTHAAKOT 762
Db 1811 KQTKVGDIAKTRD-----PALHLRIPOGALGSPRA--AAPTIIGSPPTPAGLST 1862
QY 763 GONG-PKRPATLGADGPPRPT-----PPRRTSRA-----PIEP----- 796
Db 1863 SADQFLKRP-----GSVPDPDBGLFLKLPQVPAQASQDPFGIAPVLEPRPTA 1917
QY 797 -----TPASEATGAPTRPPAPSPSAP-----PVVAPKEKEETAAAPVVEGE 841
Db 1918 PRTYPRVPSGTGAQAPRMIGASRPAGQGEHNTTPTGTRHQSTPDPFLKPF----- 1972
QY 842 EQPRAAEELAVDT-----GKAEPYKSECTEBAEGRPKGDAEAEATGALAKAEK 895
Db 1973 --RCPSLDNLAIVESPVGSGKASBPGLSP-----PPGSRK-----ALEVK 2014
QY 896 KEGSGRATTAKSGAPQ-----DSDSAT-----GSADVDEAGDKRLSLSPRSL 945
Db 2015 BEIGA-----SSPYGPNLGFVDSRSSGTLGLGELKTPVFKAP-----LTPRSQV 2063
QY 946 TPTGDPANASPOKPLDKQKORAAAPRIQYTKVNEPRRED-----AAPTQP 994
Db 2064 EPQ--SPGLGRPQEP-----PPAQALAPSPFHNDIFRGSYTDPAVQPLT 2109
QY 995 APPAPPPQN--LQPSDAPQOPGS-----SPGKSNRP-----PADKAPAAEAKLP 1043
Db 2110 PRQPPPPSCCALPERS-LPSDPFSRVVPVSPQSSQSLPTRPLPSAFAFC----- 2161
QY 1044 GPPPCWTSGLPFPVPREVIKASBPADPSAFSAPRGHPL--PLG-LHDTARPVLP 1099
Db 2162 -----PSVTPPRF-----QSPDPYS--RPSRQSRDPPAPLH--KPRPQP 2199
QY 1100 PTISNPPLISSAKHPSVLERQIGALSQGSVOLHV-PVSE-HAKAPVG-----PVT 1149
Db 2200 PEV-----AFKAGSLAHTSLGA--GFPALLPAPAGELAKVPSGQPPNVRSPT 2249
QY 1150 --MGDLPM-----DPKTLAPRSGVKQEOQLS-----RQAGPSELGVPTAOBAS 1193
Db 2250 GAFVGTSPSPKRTFPPQAVGERSLKP--VPQGLPFPAGINSHFGGPTLQKQOSTNYT 2306
QY 1194 VL-----RGALGSVNGGS--ITKGIPTSTVPS-----DSAITYGSITHTPADVLY 1239
Db 2307 VATGNHPSGSPPLGSPGSGTGESYGLSPPLPVSILPPAPADGSLPY--LSHGASQ--R 2360
QY 1240 KGTITRIIGDSSPSRLDRGRDELPGKHVLYEBSKKGHVLYEGGMSVTQSKED----- 1293
Db 2361 SGITSPVEKKEKEDPG--TGWSSILATAEI--PGTQDPGMS--GLSGTELEKQRRQRLR 2412
QY 1294 -----GSSSGPPIETAAPKRTYDMEGRGRAISSAIF-----G 1329
Db 2413 ELLIRQOIQNTLROEKETAAAAAGAVGPPGSGAERSSAPFQOLSQGTGPTPAQTODKS 2472
QY 1330 LMGRAIPRERHSPHLKEQHIRGSIQGIPIRSVYEAQEDYLRRKAKTLKREGTTPPP-- 1387

QY 337 PFEIRKORLEOERMOSR-----VCGRSGLSMSAARSEHEVSEIIDJISEOENLEKQMRQ 391
 Db 1192 LOEAPFGELELDLSRKALFAVGVGRPSFGIGTPKAKG-----DGSSEKLEPTS--Q 1241
 QY 392 LAVIPEMLYDADQORIK-----FINNGLMADPMKYKORQVMMMSSEQKETFREKFM 445
 Db 1242 KGDGDPDIADSESGLEBKADTPGEDGCVKASVPV-----SDPEK----- 1282
 QY 446 QHPKNFGLIASFLEKETAECVLYYYLTKKNENYKSLVRSYRRRGKSQOQOQOQOQOQO 505
 Db 1283 PGTPEEGMLSSDLIDISTEE----- 1302
 QY 506 QOQOQPMWRSSQOEKEKEKEKE--AEKEE--KEVENDEKDLK----- 547
 Db 1303 ---LTK--MESDLOQLFDVLSEREGHGGCTPGLSGSPPLQRPLOGLGPLYGN 1354
 QY 548 ---EKTDUTSG-----ENDDEKAVASKGRKTANSQGRKGRITREMANEASEEAIT 597
 Db 1355 LPSSSPMDSYPLGCSPPILDSNERGFPSPPEBGEPSWTSQGTTPSTPT--T 1406
 QY 598 PQOSAEIASMELNESS--RWTEE--METAK-----KGLIEGRWMSAIAIRWVG 642
 Db 1407 PTTBESGGLGSLYNQSLQWMEKDELGOLSTISPLYANINPNLKODYDPMWS----- 1459
 QY 643 SKTVSOCKNFYENYKRONLDEILQOHLKMEKENARKKKKADAAASEAAPPVVED 702
 Db 1460 ---SRCKQIMKMKRPAADKAPYLQKAK--DNBAARINKVQKQAESQ-----IN 1505
 QY 703 EEMESGCVSGNEEVEEAEALHASGNEVPRGECGPRATVNNSSPTESI PSPHTEAAKOT 762
 Db 1506 KOTKQGDIAKTDR-----PALHLRI PQPGALSGPPRA--AAPTITIGSTTTAGLST 1557
 QY 763 GQNG--KPRPATLGADOPRPGPPT-----PRRTSRA-----PIEB----- 796
 Db 1558 SMDGFLKPPA-----GSVGPDSPEGLFLKLPQVPAQAPSQDPRGLAPVLEBRPPTA 1612
 QY 797 ---TPASBATGAPTPPPAPSPSPAP-----PVPVKEKEKEETAAAPVEBGE 841
 Db 1613 PTPYPPYSPPTGAPAPQRPMLGASSRPGACQPSFHTTPGTGRHOPSTTDPLK-- 1667
 QY 842 BQKPRPAEELAVDT-----GKAEPVKSCTEBAEBGPAKGDMAEAATLEGALEK 895
 Db 1668 --RCPSLDLAVAPESGVGGKASPELSP-----PPGSESK--ALBYVK 1709
 QY 896 KEGSGSRATTAKSSGAPQ-----DSDSAT-----CSADEVDEAGGDKNLLSPSL 945
 Db 1710 EELGA-----SSPYGDPNLGFVDPSSGTHLGLELKTDPVKAP-----LTPRASQV 1758
 QY 946 TPTGDPANASPOKPLDKOLKORAAIPIQVTKNHERPRD-----AAPTKE 994
 Db 1759 EEP--SEGLGLRPOEP-----PPAQLADSPSHPIIFRPGSYTPRYAOPPLT 1804
 QY 995 AAPAPPON--LOPESDAPQOQGS-----SPRGSRSAP--PADKEAPAAEAKOXL 1043
 Db 1805 PRDQPPPPSCCALPPRS--LPSDPFRSVYVSPQSSQSLPRLRSLSEATC----- 1856
 QY 1044 GDRPCTSGLPFVPREVYKASPHADPSAFSADPPGRL--PLG-LHDTARPLYLPP 1099
 Db 1857 ---PSPVTRRF-----QSPDPYS--RPPSRPQSRDPPAPLH--KPRPQP 1894
 QY 1100 PITSNPPPLISSAKHPSVLEKQIGALSQMSVOLAH--PYSE-HAKAPVG-----PVT 1149
 Db 1895 PEV-----AFKASGLAHTSLGA--GGFPALPAGAGELHAKVPSGQPPNVPASPGT 1944
 QY 1150 ---MGLPLPM-----DPKTLAPSGVYKQOQLSP--RGQAGPPRESLGVPTAOEAS 1193
 Db 1945 GAVVGTSPKRFTRPPQAVGERSLKP--VPQGLPRRHGINHPRBGGLGKPGSTNVT 2001
 QY 1194 VL-----RGTLASVPGGS--ITKGLPSTVPS-----DSAITTREGSTTHGTPTADVLY 1239
 Db 2002 VAWGNFHPGSPGLQPSGSGTGESYGLSLRPPSVLPPAPDGLIPY--LSHGASQ--R 2055

QY 1240 KGTITRIIGEDSPSRUDRGREDSLPKGVHIVYEGKKGHVLVEGGSVMTOCSKED----- 1293
 Db 2056 SGITTSPEKREDDG--TWGSSLATAEL--PCTODPGMS--GLSQTLERQORQRLR 2107
 QY 1294 -----GRSSGPPHETAPKRTYDMMEGVRGAISSASIE-----G 1329
 Db 2108 ELLIROIQONTLROKETATAAAGAVPPGSGABSPSPAFEOUSRGOTPPAGTODKSS 2167
 QY 1330 LMGRAIPRENHSHHLEKQHHTIGSTTQGI PRSVYEAQODVLYREAKLLKREBTPPP-- 1387
 Db 2168 LVG--LPFSK-----LSGPILG--PGSF--PSDDRUSR-----PPPPAT 2200
 QY 1388 PPSRDL-----TEAYKTOA--LGPLKLPRAHEGL-----VATYKAGRSIH--EI 1428
 Db 2201 PSSMDVNSROLVGGSQAFYORALVPYPSLPLQOQOQOQLWQOQATATSRFMSABRPST 2260
 QY 1429 PREEL--RHTPELPLA-----PPRLKSGITQCTPLKYDTGASTTSKKHDVSLGSPG 1481
 Db 2261 PGPGLRQALGSLAGISTRLPGP-----GEPVGPAGPQOTELRHNVQKGLQPGG 2312
 QY 1482 RTPPPVHPLVMDARALEBACVBEISKSRPTASSGSGSIARGAPVIVELGKPPQSP 1541
 Db 2313 TPFP----- 2329
 QY 1542 TYEDHAPFAGHLPRGSPVTMRBPTPLQEGSLSSKASQDRKLTSTPREIAKSPSTVP 1601
 Db 2330 SEDH-----RLAPBGLR-----GLAVSGLPQKPSAPAP--ELNNSLHPT-- 2369
 QY 1602 EHNHPISP-----YENHLRGVSGVDLYRSHITPLAFPTSTPKIP--LDAANAAYL 1651
 Db 2370 ---PHTKGPPLPTGLELVNRPSPSTELGRPN--PLALEAGKLPCEBDELDDFAHKALED 2425
 QY 1652 PHHLAPPTYVHL--YPPVLLRGVDPDTALENROT----- 1684
 Db 2426 DEELA-----HGLGVDAKGDDELGTLENLEKNDHDLNDLNGDFDLATDPELDT 2479
 QY 1685 ---IINDYTSQOMHNATAMACORADMLRGLSPRESSIALNYAAGPRTIDLSQVPH 1739
 Db 2480 GDXKDIPREHRLVE--SANEAEAREBALLRGVEP-----GPIG----- 2515
 QY 1740 LPLVLPPTGPTAMDBLAVLPTAPQPF--SSHHSSPLSPGPGTHLTPTTSSERE 1797
 Db 2516 --PEERPP--PAADASEPRLASVLPVKKRVEEGHPS-----PCQFT----- 2555
 QY 1798 RDRDRERDREREKSLITSTTVEHAPI-----WRPG-----TEOSSGSSGSG 1842
 Db 1843 ---GGGSSSRPASHAHQHSPIBRTODALQORPSVLHNTGMGITTAVPSKPTV 1897
 QY 2598 HTAEKASFGATGCPRAH-----LTPSPISGSGSSLLEKFE----- 2634
 Db 1898 LRSTSTSPVAPRATFPFATHTCPISGTLDGVPYPTLMEVULLPKCAPRYAR--BRPADT 1955
 QY 2635 LESGALTLPGRAPAS-----GDELD--KMSLSVASLPLILELLEHEKKE-- 2679
 Db 1956 GHAFIAPKPPASGLEBPASPSKSGSEPRVVPVSGHATARTAPKMLAHHASDPRAP 2015
 QY 2680 ---LOKQOLASQLOPAQOQOQOQOQOHL-- 2712
 Db 2016 ASASDPHREKTOKSPFSIOLELNSLGNSSYSPEGVEPVSPVSSPSLTHDKGL--PRL 2074
 QY 2713 QMSLPHHGSSPSLAGSQOOL--SLGL-AVARQGLRQPLMPQOPRPHALQOARLAPMA 2768
 Db 2075 BELDSSH--LEGELARKQGVYKLGGEAHLPHLRLPESOP--SSPPLQTAFCVGHQOR 2131
 QY 2769 WVSNOGHMLSGQ-----HGQAG--LVPOSSQVLYSKKPMGTMRPMSCKMQ 2814
 Db 2132 VVTLAQHISEVI--TQDYTRHNPQOLASAPL--PARLYSPGAS-----CVYDLAR 2178
 QY 2815 QIAMAQOQLANSFPPTDLDKFALEDIIGPIAKKAVVALKGIKVAQOSIGVAPGMNQO 2874
 Db 2179 -----PPSDLYLPPPDHGAAPAGSPHSEGKSPBNKTSVLCGGEDEGIEPVSP 2227

```
QY 979 TKHNEREDAPTKRAPRPPQ-----NLQESDAPQQP-----G 1016
Db 1008 -----AVPAPPPAPPPERRKCFVLQASLRPEALEAVPRENKRAVRPAA 1057
QY 1017 SSPGKRSRP-----APPADKEAFAEAQKLP--GDPPCWTS-----GLPPFV 1057
Db 1058 SKPSTKSVPOISVGTQGGSGGK-----SGMDRPLGSSPPTALQVPLGTQLP 1113
QY 1058 P-----PREVIKASPHAPDS-----AFSYAPRGHPLPLGLHDTAPVLPRPPTI 1102
Db 1114 PASLFISQQLLPQEQOSSSEFFPTQAMAGLISPSYMP--PLPSLPLQA--PLPLQPTV 1169
QY 1103 SNP-----PLISSAKH-----PSVLEQIGALSQGMSTVL-----HV 1135
Db 1170 LHPQLHLPLQLPPLHAPADI PFOQPPFLMPCPAPSTLSGVLPLQSQFALQPLPEISHL 1229
QY 1136 PYSEHAKAPVGVPMGLPLPMDPKKLAPFSG-----VKQQLSPRQDAGPESLGVPTA 1189
Db 1230 P---PVKTSLEPLATG---PPGSSSTESSDIQLPPTPQATSPAPTSAP--LALPAC 1281
QY 1190 QEASV-----LRGTALGSV-----PGSITKGI PSTRVSDSAITYRG 1227
Db 1282 PDAMVSLVVPRIQTHMPSYGSAWYTLISQILVYQSPGSPASTALTKEEPPSSKMTV-- 1339
QY 1228 SITHTPADVLTKGTIRIIGEDSPRLDGRBDSLPKGNHYEGSKGHVLSYEGMSVT 1287
Db 1340 -----CEADV-YE-----AEPSPSISKEONRGYOTYPLRVBERKGTSLSSGILSLE 1386
QY 1288 QCS-----KEDGRSSSGPPHE-----TAAPK 1308
Db 1387 GCSSTAGSGKRVLPAGSLBLTMTQOQKRVKEBASADKELVYSCSVVLSTBERK 1446
QY 1309 RTYDMEGRVRA-----ISSASIEGLMGRALPERRHSPHHLKEQHHRIGSTQG--IPR 1361
Db 1447 KTEKPHVGGQGRSRRBAETLSSLDVSDPKELSPLSHS-----TLSHGTAPG 1494
QY 1362 SYVNAQEDYLAREKALKREGTPPRRPPRRSLTEAVYKQALGPLKRAHGLVATYKEA 1421
Db 1495 S--BALKEVAPSSK-AHRRGLP-----MSVVK 1520
QY 1422 GRSIHETPREELRTPELPLAPRLKEGSITQTPLYKDTGASTGSKHDVRLISGSPG 1481
Db 1521 -----EDPKQGTDLPLAPRPSPLSD---TSPPRAKIQEG---TDSKK-----V 1559
QY 1482 RTPEPPVH-----PLDVM-ADAPALERA--CYEESLKSRRGTASSGGGSTAR 1524
Db 1560 LQFSLHTTTVSWCYLNYIKPNHIQHADRSSVYAGWICISLYPNLPGVSTKALSLIR 1619
QY 1525 GAPVIVELGKPROSPLTYEDHGAPRPAHLPGRGSPVYMBETPLRLQESLSSSKASQDRK 1584
Db 1620 -----SKQKYSKETTYMATAP--HPEAGRLVPSNSRKPMTTEVHLNVSVSPESQK 1667
QY 1585 LTPPREIAKSPHS-----TWPEHHPHIPSEYHLLRGVSGVDLVRSHIPLAF----- 1632
Db 1668 ---DPAVKEKEKQKAEGTPTSKRGSPAVKIFE-----GG---YKSMBEYIYVGR 1715
QY 1633 -----DPTSIPIRGIPLDAAAYVYLRHL-----APNPTYPHLYPPYLIRG 1672
Db 1716 GRGRVCECGIRCKKPEMLKXHIRTHDVAVYCKCHFAFKTGNLTJKMKKASHSK 1775
QY 1673 YPDPALENRQTIINDYITSOQMHNTATAMAGRAMLGSPRESSLALVYAAGPRGI 1732
Db 1776 CQETGVLEELA---EBGTSDDLHOD-----SEQOEB-- 1804
QY 1733 DLQVPHLPVLPPTGTPATAMDLAYLTPAQPFSSRHSSPLSPGPGTHLTKPTTS 1792
Db 1805 -----AAEAVE-----EHQFSDLE-----DS 1819
QY 1793 SSGERDRDRDRDREREKSI LSTTTEVAPLWRPETEQSSGGSSGSGGSSSRPA 1852
Db 1820 DSDSDLDDEDEEBEESODELSGPCS-EAAPCLPPTLQENSPP----- 1864
```

```
QY 1853 SHSHAHQSPISP-RTODALQORPSVLANHGMKGIITAVEPSKPTVLSTSSPVPA 1911
Db 1865 -----VEGQADSDISDEVPEGGST-----SEATHLTASSCSPTSPQTQ 1903
QY 1912 TFPPTHCPRLGGLTLDGYPTLMEPVLLPKAPVAREPERADTGAFLAKPPASGLEP 1971
Db 1904 GLPRLGLAPL-----EKDMSAPSPTSBRDMS-----PSKEAGSRP 1942
QY 1972 A-----SSPKSGSEPRVLVPVSGHATARTPAKNLAPHASDPAPASASPHNEKT 2026
Db 1943 SVTRKSHLTKNDSPQOCSFAREAOASVTSTPGQWQGR-----DLGPHLCSPPBELS 1997
QY 2027 QSKPFSIQELELSLGVHSSSPGEGVPPSPVSSPSELTDKGLPKLLELDKSHLEGE 2086
Db 1998 RLTPYPI-----GREAPAGLERAT-----DTGTPKSPTRRNSLGGABS 2036
QY 2087 RPKQ--PGPVKLGEEAHLPH-----LRPLPESQPSPPILQTAGVYGHQRVVTLAQT 2139
Db 2037 PQTVLPGKMALAGPCSPADKSGGLGVPVRA-----LLQVPV--LPH-----TLLSRS 2084
QY 2140 SEVITDQYTHHHPQULSAPLPAPLYSPFGASCPLDLRRPSDLYLPPDHG-APARGSP 2198
Db 2085 PETCTSAMRKTESRSPSAG--PGPLPFPFSA-----PHDF-----HGHLPSRSEE 2128
QY 2199 -----HSEGGKRSPEPNKTSVLGG-----GBDGIPEVSP-----EGMTEPGHSR 2238
Db 2129 NLBSHLPLHSLQLSRACP--LIPIGIQVQVAPRGQPTVLPQPCAMVSGFSGGSD 2186
QY 2239 SAVVYLLYRDGEQTEPSRMG-SKSPGNTSQPPAFPSKLTESN 2279
Db 2187 TGA-----REAGERRMSPTSPSASVSPVAKVSFTLSS 2221
```

```
RESULT 17
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 27-Oct-2003
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homology
A:Reference number: Z14954; MUID:97388474; PMID:9247308
A:Accession: T03455
A:Status: preliminary; translated from GS/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PID:AA051735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: acute lymphoblastic leukemia protein, ALR type
C:Keywords: alternative splicing

Query Match 3.6%; Score 470.5; DB 2; Length 4957;
Best Local Similarity 20.3%; Pred. No. 7.1e-09;
Matches 558; Conservative 284; Mismatches 891; Indels 1019; Gaps 144;

145 LEVPSPSPPTHTPELELVPPRLSKER-----LIONMDRVDEITWVEQOISLK 194
Db 981 VKPVAFAVAP-----ELVPMKVKEPEQYFRFEGVWLTEGMALRLNLTN-----SPLH 1029
QY 195 KKQOQ-----LEERA-KPPEPEKVPVPPPIESKRLVQIYIDENRKAAMARILEG- 247
Db 1030 KRRQRGRRLGLPGAGLEGSEBPDALGPDDKKODJLTDELKKEGGEVHEKECIKLEGP 1089
QY 248 LGPQVELPLVNOPEDTQYHENIKINQAMRKL-----IIFYKRNHAKQWKQKFC 299
Db 1090 VSPDVE-PGKEETESK-----KKRKYPRGIGGFVWYRKQKSHTRTK-KQPA 1136
QY 300 QR-----YDQMLEALEKKVERIENNRRAKESKVAEYVEKQ 336
Db 1137 QAEVLSDGQPDDEVIPADLPAGAVGSLAEQDEKKQ-----RRGRKRSKLEGMFPAY 1191
```

QY 1791 TSSSERERDRDRDRERERKSLTSTTVEHAPITWRPQSSGSSGSGGSSSR 1850
Db 2222 PALSE-----GSSSEATTPVSSVAREFP---PGLF-----AAQSAEGISGKE 2263
QY 1851 PASHGHAHSPISPTTODALQORPSVLNHTMGKIIITAVEPSKPTVLASTSTSSVPBA 1910
Db 2264 SAASH-LMDLTPLSAPASLIDLAPA-----PA 2290
QY 1911 AAFPATHTCGLGTIDGVYPTLME-----PYLLPK-----EAPR-----V 1945
Db 2291 PAPAPAPGLP-GDLGDTLPCTPCTGELTKPSPFLSPSGHEANGPESTSLNPGFVT 2349
QY 1946 AAPERPRADTGHAF-----LAKPARSGLEPASPSPKSGEPRLPVVSGHATI 1994
Db 2350 AAEEBEAEPAPAMRGSGWPEGAERSRDTLLSSEQPLRPKSGSGC-----PCGLSSEV 2405
QY 1995 ARTP---AKNTAPH--HASP---DPPAPASADPHREKTSKPSIOLEL-----RS 2040
Db 2406 EAGPGCATDPRPHCGELSPSFLANPLPSP-----TDDSDLSTEARLAGKGRRR 2456
QY 2041 LGYHSSVSP---EGVEVSPVSSPLTHDKGLPKHLELDKSLHEGLRPKQGP---P 2093
Db 2457 VGRPATGCPMADETPPTSASDSSQSOSDSDVPEETEECPSTAAALDSDEGDPLP 2516
QY 2094 V-KLGG-----EAAHLPHLRPLPESQSPSPLQTAPGVKGQVVTLLAQHISEVITQ 2145
Db 2517 VDKAGVGSTHHRPGHDPPTPLPDRPSP-----RP 2550
QY 2146 DYTTHHPQOLSAFLPAPLYSPFGASCYVLDLRPPSDLYLPPDHGAPARSPHSEG--- 2202
Db 2551 DVCMADEPGLS-----ESGRVERLEKGR-----PGRARAGRAKAPSPARLL 2593
QY 2203 ---GKSPSPMNTSVLGGGEGDIEPVSPPEGMT-----EPGH---SRSAVYPLVRDQGOT 2252
Db 2594 DIRKRSPTPGKPYDRISRTVPRRSTPSOVTSAEEKDGHSPMSKGLV-----NGIKA 2647
QY 2253 EPSRMGSKSPGNTSQPPAFPSKLTESNAGMYSKKQKQIINKLINTH-----NENEP 2302
Db 2648 GSTALGSK---GSGSPYVVDLAYIPNHCSGKTDQDFRRVRASYYVVGNDP 2698

RESULT 16

T42717
DNA-binding protein Rc - mouse
N:Alternate names: Ig kappa chain gene enhancer Recognition component
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T42717
R:Wu, L.C.; Liu, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.
Genomic 35, 415-424, 1996
A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for
ew family of large transcriptional proteins.
A:Reference number: 222238; MUID:97001141; PMID:8812474
A:Accession: T42717
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2282 <MWL>
A:Cross-references: EMBL:L46815; NID:91377885; PID:91377886; PIDN:AA040884.1
A:Experimental source: strain BALB/c; clone TL; thymocyte, brain
C:Genetics:
A:Gene: Rc
C:Function:
A:Description: binds V(D)J recombination signal sequence and kappa B motif
C:Superfamily: HIV-EBP enhancer-binding protein
C:Keywords: DNA recombination; transcription factor

Query Match 3.6%; Score 477; DB 2; Length 2282;
Best local similarity 19.3%; Pred.No. 1.8e-09;
Matches 518; Conservative 288; Mismatches 900; Indels 976; Gaps 127;
QY 7 LVATWATERPRYPHSGSYVQIARHTTVDGLLEYGHHSDVASHSPSSITIQPORRRP 66

Db 107 LEGSTWQVDMRPGSGSFVA---PGSHPGQL-----PSHAS-----ILPEELP 151
QY 67 SLSEFQGNRRSOEHLRPPESHVYLPGLKSEMEFIESKRP-----LELLDPLLRP 120
Db 152 GIPKVFVP---RPSQVSLKPAEBAHKKE-----RKPKQPKXYTCQYCSRPQAKP 197
QY 121 SPLA-----TGO-----PAGESDLTKDSLTGKLEPVSPSPPH-----TDPELEL 162
Db 198 SVLQKHTSHGTEGPPYCGPGCFSPKTSNLYKFR-----KSAHAIKGLAGSSGSEW 251
QY 163 VPRILSKEELLQNDRDVREITWEEQOISKLKKQOOLEEBAAPPEPEKVPSPPIESK 222
Db 252 YPPGL-----EMERIPG-----EEPEEPTGEGSTSEBTGAASPGSTDLKPK---K 296
QY 223 HSLVQIITD---ENRKAEAARILLEGIGOV-----ELPLYNOSPDTROYHEN 269
Db 297 HPLLSSLSYSGSHSGSSQERCSLSQSGSTGSLIEDPAPFAEASHPHLSKPEDHTTKOK 356
QY 270 IKINQAMRKLLILFKKRNHARKQKQKFCORYDQLMLEALEKYERIENPRRARESKV 329
Db 357 LALNLSERKLLI---EQTLPSQSGKSTSGTFSNSESABQOV---SPPTNAK--- 405
QY 330 REYYEKQPEIRKQRELQERWQSGSGLSMAASHEVESEITDGLSEQENLEKOM 389
Db 406 -SYAEIIFKGC-----GRIGRTSMIASTSTQ----- 431
QY 390 RQLAVIPMLYDADQSKIKFTINMGMLADPMKYKQKQVNNMWSDEKETFREKFMQPK 449
Db 432 ---PLPSEDEPRLV---PLSVP-----RTOVIEH--- 457
QY 450 NFGILASFLEKRTYAEVCVLYYLTKNENYK---SLVRSYVRGKSGQOQOQOQOQOQ 505
Db 458 ---ITKLI---TINEAVVD---TSEIDSVKPRRSILTRKSSVESPSS----- 496
QY 506 QOQOQPMRPSQOEKE 550
Db 497 ---LYRDSLSSHGKTKQEOQLSLQHPPESTHVPPLRSHSPMSAACTISTHHTF 550
QY 551 -DTSSEENDKEKAVASGKGTANSQCR---RKGRIITRSMANANSEATTPQOSELAS 606
Db 551 RGSYSFDDHADPEVPBRNTPVFTSHPRMLKRAAAIELPLGGEYSSE---PPSSKDPT 607
QY 607 MELNESSRWTEBEM-ETAKKGLLEHGNMGAIAAMVGSKTIVSQCKNFYFNKKRONLD-- 663
Db 608 SKSDDEBEPKESDLTKTKTKGPKTKGANY-----ECTICGARYKKARDYEAH 654
QY 664 -----EILQGHKL---MKERK-----NARRKK----- 684
Db 655 KKYCYSELQITKAHSVGAHEVEKTOABRPWQMMHYKGLATLELPLRKRRKEKSLGDE 714
QY 685 -----KAPAASEEAAPPVVEDEMEASGVSGNEEWEVEA 721
Db 715 EEPAPACPPGSTAHNRPGLSTKSPAEASKSA---PSLEDPASSPGLPSOE----- 764
QY 722 EALHASGENEVRGE-CSGPATV---NNSSDPEISPSHTEAAKOTGONGPAPATLGADG 778
Db 765 -----LQNGQRRRQEQPKKFTVIQHTSSPEKSP-----PEQSGLEEDKP 806
QY 779 P---PGPPTPPRTSAPALIEPTPASBATGAPPPPPAPSPSAPPPVPEKEKEEETAAAP 835
Db 807 PAQSSPPPPAPHGRSAHSLQPLVRQ-----PNIQVEILVTEBDRPTEPER 855
QY 836 PVEGEQKPPAAEELAVDTGKAEEVVSCEETEE--AEGCPAKGP---AEEAEEATAEGA 890
Db 856 PPEPEK-----TEEFQWPKOTUQAQLPAELKPPKKRLRLAEMAQSSGSES 903
QY 891 LKAEEKGSGRATTAKSSGAPODSSATCSADVEDVBAEGDKRNLSPRP----- 942
Db 904 FESSVY---LSNRPGQESSISLSSGSRASFPREDHGAE-----APGSPDTRSKT 952
QY 943 ---SLTPTGDD---BANASPKR-----LDLKQLKORAAAIPTPIOV 978
Db 953 LGSMLTLVPSHHPHAREMRSASQSGPNVPHSSHMTETRSKSFYDGSLSPTGPSTL----- 1007


```

QY 1192 ASVLRG-----TALGSPVGSITK-----GI-----PSTRV----- 1217
Db 783 DSLAMGRREQDPTLAESSDDFGSLRHHYVEDSSSGSLPLPQPPADAMTDDEFMRQ 842
QY 1218 -----PSDSAITRGSIYTHGTPADVLYKGTITRIIGEDSPRLDRGRD-SLP 1264
Db 843 ILEMSAEEDNLEEDDTAVSGRLAGHGAQK-----ASARPPRESSQSVALP 889
QY 1265 KGHVITYEKKGH--VLSYEGMSVYQCKEDG----- 1294
Db 890 KRRLPHNATYVEILLSEEGPAEPTDGLQGLRFRKTIGLNSTGRLMSTSLDGLQGSDE 949
QY 1295 -----RSSGPP-----HETAPKRTYDMEGRVRAISASISIGLMRAIIP 1337
Db 950 NLDRPEELMESTLCSPEDRSGEHSSTLPASTPSTYGTSTSLSEED---SDSPS 1006
QY 1338 ERHSPHLKEQ---HNRGSIYTOGIPRSYVEAOBYLRRKAKLLKREGTPPPPPSDL 1393
Db 1007 RRRLLEAKQOKRKAHRSHGPL--LPTIEDSEEBEILREBEILLREQEKREVEVQQRIR 1063
QY 1394 TEAYKTQ-----ALGPLKLAHEGLVATVKEAGSHIEITREELRTHPE 1438
Db 1064 STARTTRDKELRAQRERRSKTPPSNISP-----IEDAS-----PTEELRQAAE 1109
QY 1439 LPLAPRLKEGSIYTOGTPDKTDTGASTGSKKHVRSILGSPGTFPPVHPPLDVADARA 1498
Db 1110 M-----EELHSSCSGYSP-----SPSLDSEMTLDGGPTRIYKSGSEYNLPAM-- 1154
QY 1499 LERACYEESLKRPGTASSSGGSARGAPVIVPELG---KRPQSDLYTEHGAAPPAGHL 1555
Db 1155 -----SLCSPTETPGSGSTTPSGGRPLKSAEAYEDMMRKALLORQOQOAGARGP 1206
QY 1556 RGSPTWKEPRLPQEGSL-----SSKASDRKLTSPRRIASPHSTVEHHHP 1606
Db 1207 HGGP---SQPTGPRQSGFEYQDTLDHDYGGASQP--AADGTPAGLG---ATV----- 1252
QY 1607 PISPYEHLRGVSGVDLYR--SHIPLAFDPTSIPIGILPDAAAYLPRHLAPNTYPL 1664
Db 1253 -----YEELIQTOSIARMKQASSRLATTEDEKKKQOLMESAYMDMKONGSGYLGT 1308
QY 1665 YPPLYLR--GYP-DTAALENQTIINDYITSQOMHNTATAMAQADMLRGLSPRESSLA 1721
Db 1309 SPTQLAAVSPFTSTSSDSSGGRVLPDVRTQ----- 1340
QY 1722 LNYAAGPGGIIDLSQVPHLPVLVPT-----PCTATAMDLAIYPTAPQPPSSHS 1773
Db 1341 -HFAKEPQEPILKLHSSPASPSLASKVEMTFSQGGTPEAT-----TMAPCPASLP 1390
QY 1774 SSPSPGPGTHLTKTTTSSSERERDRERDRERKSLITSTTVEHAPIMRPGTEQ 1833
Db 1391 RGYMTAPGER--SPSTSTHSHYQOPTTANYGQTEB-----LPHADSGPAGSGR 1440
QY 1834 SSSGSSGSGGGG-----GSSS--RPASHAHQHP--ISPR-TODALQ 1873
Db 1441 ASREKPLASGGGEVGPQPSRGVSYFTSSPSLSTSESTSESTPSSLGPRATAEFTQ 1500
QY 1874 RPSVLHNTMGKGIITAVEPSKPTVLKSTSTSPVPAATTPPATTCPLCGTLDGYPTLM 1933
Db 1501 TPS-----LTPSS--DIPRSVGTSPVMAQGTQTP--HRPSTPRILVMOQSSQE 1544
QY 1934 EPLV--LPEKAPVAPRPERPRADT-----GHAF-LAKPPRSGL--EPASP 1975
Db 1545 APVMTITLASDASSQTRVNHASASTSPICSPPDQSPASHSTYQTPPPASQMPSEPAQP 1604
QY 1976 SKGSEPRPLVPVSGHAT--ARTPAKULA-----PHASDPD--DAPPASADPHRE 2024
Db 1605 ---GPRPARSAVGSPRLALYGMGALPAENISLCRISSVPGTSRVERGPRGTAVVDLT 1661
QY 2025 KTQSKXPFSI--QELFLRSLGHNGSSYSPBGVEPVVSPSLTHDKGLPKHLELDKSH- 2081
Db 1662 AVKPPPIITLQGMULTSLAVEARKYG--LALDPVPGROSTAVQPVIVINLMAQEQHT 1717
QY 2082 -----LBEELARPKQGPVKLGGEAHLPHLRLPESQSSSPL----- 2119

```

```

Db 1718 FLATATTVSITMASSVLMQAQOKQPVVYGD-----PFGRLDFGQGGSGPVCGLAQVQVEQ 1772
QY 2120 -LQTAAP---GVKQGRVYTLAQH-----ISVITQDYRHHQQLASAPLAPLY 2164
Db 1773 AVQTAAPYKGRGPRREAKFARVYMLPNQVTLAARDLITITQMTAOSVSLKGPVPE-- 1829
QY 2165 SFGASCPVLDLRBPSPDLVLPDPDHGAPARGSPHSEGGKSPPEPNKTSVILGGE----- 2219
Db 1830 -PGAE-----PHRATPALRAHALPQTKPH---TVVQMEGAAGT 1867
QY 2220 -----DGIEPVS-----PEGWTE---P 2234
Db 1868 VTTLLPEEPAGALDLTGMRPESRLACCDMAVKPPFGSSCTGTPHPAPAPKSVTDALP 1927
QY 2235 GHSRSAYVPLLYRQGEQTEP---SRMGSKSGQNTSQ-----PPAFSLUTEEN--SA 2281
Db 1928 GQSSGPPFYS--PRPEPEPEPLTPFAQGVGPGPEHQRYPQGLPGRLYSSMSDNTLAE 1985
QY 2282 MYKSKKQRIKKMLNHN-----EPEYNI 2306
Db 1986 GLNTHAQRIQGLFQGGHDSAVDLSLKHYSYLGFPADGRYLGQGLQYGSFTDLRHPTLL 2045
QY 2307 SQP-----GTEIFNMPAIT----- 2320
Db 2046 SHPLMPRVSSVSNISYDHRVGRPGDAVGFQEASLAQYSATTAREISHMCAALNSMDQYG 2105
QY 2321 -----GTGLMYRSCAVOEHA---STNGLERITIKALMGK--YQMEESPPLSANA 2367
Db 2106 GRHGGSGGPDLPVPOQ---HGGLNAPQGL--ASLSGLGNPTYEGQPSGNLQ-- 2159
QY 2368 FNPLNASAS-----LPAAMPITAADGREDHTLTS----- 2397
Db 2160 YCPAASQGTAVRQLPSTATYRAADGMYSITINPIAATLPTTOPASVLRPMVRGMYR 2219
QY 2398 ---GGGKAKVSGRSPSRKXSP--APGLASG-----DRPVSYSVHSEGCNRT 2443
Db 2220 PYGSG---VTAVPLTSLTRVPMIAPRVLPBAGLYRPAPSRPIASTI-----P 2267
QY 2444 PLTNRWEDRP-----SSAGSTPP-----YNPLMRQAGVMASPPPG 2483
Db 2268 PABGPVILGKPAAKASAGGPPPELPAGAREBPLSTTAPPAVIAKAPVAQAAPAPPG 2327
QY 2484 -LPA-----GSGPLAGP 2494
Db 2328 QKPADAAAGSGSVLGRP 2346

```

RESULT 15
 A43359
 microtubule-associated protein MAP1A - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
 C:Accession: A43359, S22108
 R:Langkof, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
 J. Biol. Chem. 267, 16561-16566, 1992
 A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messeng
 A:Reference number: A43359; MUID:92355629; PMID:1379599
 A:Accession: A43359
 A:Molecule type: mRNA
 A:Residues: 1-2774 <L>
 A:Cross-references: GB:M83196; NID:g205537; PID:AA848069.1; PID:g205538
 A>Note: sequence extracted from NCBI Backbone (NCBI:111039, NCBI:111040)
 R:Citavchik, A.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S22108
 A:Accession: S22108
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 73-364, 'NLR', '370', 'QKN', '374', 'PSFKGL', '381-751', 'RSMMSQMAARR', '764', 'D', '766', 'LR
 'WIKNMGPQPSQSP', '851', 'V', '853', 'NSL', '855', 'LPHRWLRN', '865', 'W', '867', 'HSQLPDGDG', '877', 'Q', '879',
 A:Cross-references: EMBL:X66840
 A:Experimental source: strain Sprague Dawley

Db 1465 HSNPA-----GIGQALQSSRGSGSVTAPGGHPR-HKRGPPQAGG 1504
Qy 1957 HAFKAPPAR-----SGLEPASPSKSGSEPRPVVPSYGHATTARTPAKNLADHNAAP 2009
Db 1505 PS--PRPTRYEPORVNSGL-----SSDPHEEBGPMVGVG-----TPRPSAG--VSP 1550
Qy 2010 DPAPASADPHREKTSKPSIOLEL-----RSLYGHSSSVSPGCVAPVSVSLSLTH 2066
Db 1551 FPP-----KRRERPRKPELLOBESLPPPHSGGFGAS--KPDG--PGQASRDTG 1597
Qy 2067 DKGLPKL-----EELDKSHLEGLRKQDPVKYLGGEAALPHLEPLP--ESOPS 2115
Db 1598 TEALPHIMNLHTATSKSYRPSMEPMMEPLSPEDVAGTEMGSDSVDLSGDSQVS 1657
Qy 2116 SSPLLQ-TAP--GVKGHQRVVTLAGHISEVITQDTRHNPOL--SAPLPA-PLYSPPG 2168
Db 1658 SGPCGQRSSPDGGLG-----AAGPPKRPFGSSPLNAVPCGEPG 1698
Qy 2169 ASCPVLDRRPPSDLYLPPPHGAPARGPSHSEGGKSPPEPKTISVLGGEDGIEPVSPP 2228
Db 1699 SEPP--RRPP-----PA--PHDGKPELP--REOPLPPGPIGTERSGRT 1736
Qy 2229 EGMTEPGHSRAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFSKLTESNSAMVSKKQ 2288
Db 1737 DRATEPGDIR-----PEHRGPPVQF----- 1757
Qy 2289 EINKKLNTHNNEPEYINISQFTELFNNPALTGTGLMYRSGAOVEMASTMGLAIIKK 2348
Db 1758 ----GTSDDXS-----DLRLVVGDSL--KAEKE-----LTSYTE 1786
Qy 2349 ALMGKYDOWESSPLSANAFNPILNASASLPAAMPITADGSRDHTLTPSGGGKAKVSGR 2408
Db 1787 ALPVRSD--WELLPSAASA--EPQSKNLDGHCVPSPSSGGRLPEVYGSAG----- 1837
Qy 2409 PSSRKAKSPAGLASGDRPPSVSVHSGDCNRRTPLTNRWEDRPPSSAGSTPF 2462
Db 1838 PSSSGISGSGHGLS-----ITSQWRRLRPGPRLHPY 1869

RESULT 13
T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Treisman, J.B.; Luk, A.; Rubin, G.M.; Heberlein, U.
Submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TR>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AA06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query March 3.8%; Score 502.5; DB 2; Length 2715;
Best Local Similarity 20.9%; Pred. No. 2.8e-10;
Matches 450; Conservative 189; Mismatches 752; Indels 763; Gaps 111;
Qy 670 KLKKEKERNARKKKKAPAA-----ASEBAFPVVEDEMEASGVSGNEEWEAEAA 723
Db 4 KIKSPQTQOQOQGGAPAPAAATPPSAGARGAATPPT----- 39
Qy 724 LHASGNEVPRGCGPATVNNSSDTESLPS-----PHTBAKDTGQNGPKPATLGA 775
Db 40 -----SGPPTPNNSNNNGSDPSIOOQONVAFHRYGA-----PPPPGS--G 77

Qy 776 DGPPPG-----PTPRRTSR-----APTEPPASATGAP----- 806
Db 78 FCGRRGPRRAAVMTYTHNLHQOQOQHPRPHMQOQHNGGRRAPRRPGCAPEHAPGVKEYET 137
Qy 807 -TPPPAP-----PSPGAPPVVKEKEEETAAPVEE--GEE 842
Db 138 HLPRPHPRAYGRVHADPNMDPRYGOGLPGGKPR--QOQOQHPRQOQPGGSP 193
Qy 843 QKPRAEELAVDTGKAEPEVKSSETEAEBSR-----AKKDAEAAEATLGLKAEKK 896
Db 194 NRPPQORYIPQRPQGPPTPLNSLLOSSNPPRPHRYANTYDQOAAASAAAAAQOQ 253
Qy 897 EGSGRATTAKSGAPQSDSATCSADVEDEAGGDKN-----RLSPR----- 941
Db 254 QAG--GPPRGHGPRPPQHPS-----PTGGQGGAPRRPRPISPOLGSSQOYRT 301
Qy 942 -PSLLPTGD--PRAN-----ASPQKELDKQKORAAAI-----PPIQVTK 980
Db 302 PPTNTSRGQSPYPYPAHQNGSVSPSPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 361
Qy 981 VHEPRREDAATKAPAP--PPQNL-----QPSDAPQOQSSSRGK 1022
Db 362 --QPQOQNTPTSQSYSPYRYPPTPGLPAGSNHRTAYSTHOYPERNRPPGSSSPSG 419
Qy 1023 SRSPAPRADKFAFAEAKQKLPDPPCWTSGLPFPVPPREVLIKASP--HAPDSAP--SYAP 1080
Db 420 SGHPLPASP--HYPLQOQDPPPHVSAGP--PP--SSSGHAPSPSPQOASP 471
Qy 1081 GHPLPL-----GLND--TAPVLPPTLISNPPLISSAKHPSVLERQIGALS 1126
Db 472 SPQELIQNSNDSSGSAHSGMGSPGTFNPPQVNRPTSPGSS-----GSRG 522
Qy 1127 QGMSVOLHPVSEAK-----APV--GPVTMGLPLMDKCLAPFSGVQKQOQLSPGQA 1178
Db 523 MSPAVQNHPISRPASQSSSGGPMQOPVYAGGPPMPRPPGWP--GAPPOQOQOQOQ 581
Qy 1179 GPPSLGVPTAOEASVLKGTALGSPGSGITKGLPSTRVSDSATVYRGSITHTPADVL 1238
Db 582 SNGASSASNSPQQT-----PPRAPPPQGNANNAATPPPP--QGAAGGYRPPPH 629
Qy 1239 YKGTITRIIGEDSPRLDGRGEDSLPKGHVLYEKKGHVLSYEGMSVTQSKEDGRSS 1298
Db 630 MHGYSK--MGQPGGSPAQGYRPPQOQ-----QYRGNVPRPPQYRPAVAT 674
Qy 1299 GPHETAAPKTYTMMGGRGKRAISASISGLMGRALPPE--RHSPPHL--KEQHIRG 1353
Db 675 GPP-----PPTSQAGAGANSMPSGAAGGYRPGRMNHTGOYRPPQWPPSPQOTVPG 729
Qy 1354 SITGIPRSYVEAQEDVLYREBAKLLKREGTPPPPPSPRDLTEAVKTQALGPLKLPANEH 1413
Db 730 ----GAGCGAMVGNH-----VQKGTIPPPVVG-----GP-----PPQG 760
Qy 1414 LVATVKEAGSISHEIPREELRHTPELAPRLKESITGCTPLKYDTGASTGSKKHV 1473
Db 761 ----SSGPRPLNYL--KQHLQHKGYGSGSPTP--PQG--PQGYNG--PTGM----- 800
Qy 1474 RSLIGSGRTIPPH-----PLDMAADALRACVTEESLKRPPGTASSGG 1520
Db 801 ----HPGMPMGPRPHNGPRPHGPTNMPPTSTPQSMQO-----GGQPOGQAGSG 847
Qy 1521 SIARGAVIYBELCKPRQSPLYEDHGAPFAGHLPRSP-----VTWREP 1565
Db 848 -----PESGGRPH--ISQDNKISSG--PTGAAGMAVTSVYTTGPDGTSNDEV 892
Qy 1566 TPRLQEGSL--SKASQDRKLTSTPREIASP--HSTVPEHNPPI-----SPYENL 1614
Db 893 S--QOSTLNASASGSDPQCTTPKSKRNDPVSQSHLAPRSPSPHVVVHPPGGPEEY 949
Qy 1615 LRG-----VSGVDLYRSHIPLAFDP--TSIRGCLPLDAAAAYULPRHIANPTYPHLY 1665
Db 950 DMSPPNMPRPASPOVFNVHVPQOEFRTITTTTKKSDLSCLXY--EMDNDPDR----- 1002
Qy 1666 PPLYIRGYPD--TAALENRQTIINDYITSQOMHNHTA--TAMQBRADMRL-----G 1712

Db 5721 -----YEKK-----QAVHKEKVFIESFEPEYDELEVEYTEPFEQPYEEFDE 5764
 Qy 263 TROHNENKINOAMKULI-----LYFKR---NHAKOMOKFCOR----- 301
 Db 5765 D---YEEIKVE---AKKEVHEMEBEDFEGOEYEREEGDEGEHEMEBAVQERFVIOVQ 5818
 Qy 302 ---YOULME-----ALEKK-----VERIENPRRRARAE---SKVREYEQ 336
 Db 5819 KEVIESEHERKVPAPVEKKAPPPPKVILKKVILEKTESRAMEEKQVTKVPEVSKKI 5878
 Qy 337 FPEIRKOELOERM-----OSRVGORGSLG---MSARSEHEVS---EIIDGSEON 384
 Db 5879 VPQKSRTPVOEVEILEVKVPAVHTKQVISEKMFASHTEEEVSUTVPEVKEI.VTEBEK 5938
 Qy 385 LEKQKOLAVIPMLYDADQ---ORIKFINNGMLADPMKY---KOROVMMMSOE 436
 Db 5939 IHVAVSKVEPPPKVPELPEKPAPEEVAVPVPI-----PKVPEPAPKVPPEPKKVPPEE 5992
 Qy 437 KETP---REKTMQHPKNGFLIASFLERKTVAACVLYLTKKNENYKSLVRSYRRBKS 493
 Db 5993 KKPVPVPPKKEPAPK-----VPPEVKPVEBEKI.PVPAKKGKAPPAKVPEV---QKGV 6045
 Qy 494 QOOQ-----QOOQ-----QOOOQOOQPMRSSOE-----EKDEKEKEKE 528
 Db 6046 TEBKITTVTOREESPAPVPELPPKKVPEBERKVPKEEVPPEPKVPALPKKVPBEKV 6105
 Qy 529 A-----EKEEKPEYENDKEDLKEKTDGSDGENDKEKAVASGRKTAQSGRRKGIT 583
 Db 6106 AVPVPAKKAAPPRAVEVSKTVV-----BEKRFVAEEKLSFAVQO---RVEVT 6150
 Qy 584 RSMANEAENSEAIPQOSAEIASMET.-NESSRMTSEEMETAKGILEHRNM-----SA 636
 Db 6151 R---HEVSAEHEMSTISEEBEGVISIVYREEREEBEELVTEYEWEEPEEYVVEEKIHI 6207
 Qy 637 IARWGSK---TVSOCKNFYFNKYKRONLDE-----ILOQHKLM---EKER 677
 Db 6208 ISKREAPAPVTEBOEKI.VLKPKIPAKIEBPAPKVPBEAPKKI.VPEKVPAPVPEKKEK 6267
 Qy 678 NARKKKKAPAAASBEAPFP--VVEDEMEASGVSGNEEVEEALHAASGNEVPRSEC 736
 Db 6368 VPPPVPEEPKPKVPEKVPKPKVIMKEBPPLAKVTKEKMQITOEKVLVA---VTKKEA 6323
 Qy 737 SGPAIVNNSSDTESIPSPHTEAAKDOTGONGPKPPATLG-----ADGPPGPPT 784
 Db 6324 PPKARV---PEEPKRAVPEEKVLKIKKRESEBPAPKVTPEFKKVVKEKVSIBAPKPEQ 6380
 Qy 785 PPRRTS---RA-----PIEPT----- 797
 Db 6381 PIKEVTIMEKERAVTLEEBAVSVQREBEVEYEYDYKPEPEYPTTEYDQVEYERE 6440
 Qy 798 -----PASEATGAPTP--PAPPSAPAPVPVPEKEKEEETAAA-----P 835
 Db 6441 YERYEEHEEYITEPEKPIPVKPVPEEPVTPKAPPAVILKAVVEEVPVPPIPKKPP 6500
 Qy 836 PVEGEEOQPPAAEELAVDTGKAEBPVSECTSEABEGPAGKADAAEATAAGALKAEK 895
 Db 6501 PPKVPEBEKKVPEEKIHISTIKRE---KEQVTEPAKVPMPK-----KVALEEKVPVPR 6552
 Qy 896 KEGSGGATTAKSSGAPODSDSATCSADP---VDEAEGDKNRLISPRSLT----- 946
 Db 6553 KE-----VAPRVAVEVPKELEPEEVAFEEEVVTHVEEYLVBEEBEYIHEEBEFTTEEBV 6608
 Qy 947 ---PTGPRANASP-----QKPLDKOLKQKRAAATPPIQVTVHPRPRDAPTPAPAP 999
 Db 6609 PVIVKQVEVPKRVPEEKVPVPPKKKG---APPAKVPPEVPKPEKVPVL.PKKEKP 6664
 Qy 1000 PPOMLQESDAPQOPGSSPRGKSRPAP--PADKEAFAAEAOXLPGDPPCWTSGLPVPV 1058
 Db 6665 PP-----AKVPEVP--KKVPBEKVVPVPAKVEAPPAKVPPEVK-----PVP 6706
 Qy 1059 PREVIKASPHAPDSDASVAPPGH--PLPLGLHDTARVULPRPTISNPPLISSAKHPS 1116

Db 6707 EKVVPAPKAVE-----APPAKVPPEVKKLIPEBEKPTVPKKAPEAPPKVPKKEBP 6759
 Qy 1117 VLEROIGAISQMGVQULHPVISEHA--KAPGVPTMGIP-----LPMDPKLAPPSGV-- 1167
 Db 6760 -----VPPVVALPOEEEVLPEEEIVPEEEVLPPEEEVLPPEEEVLPPEEEVLP 6807
 Qy 1168 KOEOLSPRGQAGPPESLGVPFAOEASVLRGALASVPGSITTKGI--PESTRVPS-----DS 1221
 Db 6808 EEBEIPPEEEVVPPEEVPPEE-----FVEEEVLPPEKVPVPAPAPPELTK 6857
 Qy 1222 AITYGSI-----THGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVITYEGKKGHV 1277
 Db 6858 KYTEKVVIPKKEAPPAKV-----PEVPKVEEK-----IILPEKEEV 6897
 Qy 1278 LSYEGASVTOCSKEDGSSGCPHETAPAKRTYTDMEGRGIAISSASIGLMGRALP 1337
 Db 6898 LPVE-----VTEPEPEEPISEBEIPEEPPSIEVEEVAPRPRVAYIKAVPEA--PTVPK 6951
 Qy 1338 ERHSPHILKEQHNRGISTOGIPRSVYEAQEDVLRREAKLLKRGTPPPPPSRLTEAY 1397
 Db 6952 KVEAP-----PAKYSKIPBEKVVPVQ--KKEAPPAKVPEVPKVPBEKVLPV-- 6998
 Qy 1398 KTOALGPLKLPAHGLVA-----TVKE-----AGRSIHETPRELRTPELPLAPR 1444
 Db 6999 KKEAVPPAKGRTVLEKVSVAFRQEVVKELELVVEABVEEILPEEEBFEVE-----E 7053
 Qy 1445 PLKGSITIGTPLYDVGASTGSKGDVNSLJSSPGRTFPPVPLD---VMADARALE 1500
 Db 7054 YFEEGERF-----HEVEFILTBOHRVEEHRVEKVRHIVFEAE 7094
 Qy 1501 RACEESLSKSPGTASSGGSIGARGAP--VVPPELGKPROSP-----LTVEDGAPFAG 1552
 Db 7095 VEVEKRP--KAPKPEPEISEKILPPKKTPTKVPPEKPEPAKVPEVKI.VAEKVRV--- 7150
 Qy 1553 HLPKGSVPYTMPEPTPLDOEGLSSSSKASQDKLSTPREIAKSHSVPEHHHPHISYE 1612
 Db 7151 BEPVPPTKVEVLPPEKVEVPEKVPAPPAKPPAPPKVPEAEKVEVPEKV--FVPPPK 7209
 Qy 1613 HILRGVGVULYRHHIPLAPRPTSIPRGIPLDAAAAYLPHILANPPTYPLPYLIRG 1672
 Db 7210 -----KEVP---PTKVPPE--VPPAAVPEKKVPALPPEESP---PPEVFE 7248
 Qy 1673 YPDTALENRQTIINDYITSQMHNTATAMAQADMLRGLSPRESSIA-----LN 1723
 Db 7249 ----- 7267
 Qy 1724 YAAGPRGIIDLSQVPH-----LPVLV-----PTGTPATAMDRLAYL 1761
 Db 7268 VPEVPKEVPEKKVPAPAPPKPEVTPVKVPEAPKVEVPPEKKVPVPPKKEVPPTKVPEV 7327
 Qy 1762 PTAQPPSSRHSSPPLSPGPTHLTKPTTSSSREBRDRERDREREKSLITSTTV 1821
 Db 7328 PKAVPEKKVPEALPPEESP---PEVPEEPEEVALPEEPAVVEPEEPAAPQVTVP 7383
 Qy 1822 EHAPIMPTGEOSSGSSGSSGSSSPASHAHQHSPIPSRTODALQOORSVLANT 1881
 Db 7384 PKNPV-----PEKAPAVAKKRELPV-----KPEVPEK 7415
 Qy 1882 GMKGITAVESKP-----TVLRSTSSVPRPAATFPATPCPLGGLDGVY 1929
 Db 7416 VPEKKVPLVVPKPEADPAKVPEVKEVPEKKVAVPEKPEV--PPAK-----VP 7463
 Qy 1930 PTLMEPVLLKPEARVAPRPRPADTGHAFILAPPASGLEPASPSPK-----GSPRP 1983
 Db 7464 EVPKKPVL--EEKPAVPEPRASPPPEV--EKEPEIABEETIAPBEKVPVVAEEEP 7519
 Qy 1984 LVPEVSGHATIAIPAPNL-----APHASPPD-----PAPASADP 2021
 Db 7520 EVPP-----PAVPEPKKIIEKKVPVITKKEAPAPPKPEBEKYLEKKLAPRRPPPPAP 7575
 Qy 2022 HREKTOGKPSI-----QELELSLGYHGS-----SYSPEGVAPVS 2057
 Db 7576 KPEOVKEKIFOLKAI.PKKKVPENPQVPEKVELPPLKVPGGEKKVRLLPKPEKPEKEEV 7635

```

QY 1339 -RHSPHHLKEQHHI-----RGSITGGIPRSYVEAOE-DYLREAKLT----- 1378
Db 1855 RSHAPRGKHHYVWQGBEAGTAVTLLPEEPAGALDLTGMRPEGLACDMVYKPF 1914
QY 1379 --KKEGT--PPPPPSRDLTE-AVKTQALGPLKKAHEGLAVTAKGASIHIEI PREEL 1433
Db 1915 GSSCGTFHPAPASAPDKSVTDJALPGQSSGPF----- 1946
QY 1434 RHTBELPLAREPL-----KEGSITQGTPLKY-----DTGASTTSGKHQDVR 1474
Db 1947 -YSRDEPPEPPLTFRTOGVVGPGRHEQRPY PQLPGRLYSMSDTNLAAGANTYAOQR 2005
QY 1475 --SLIGSGRTFPVHPLDVADALERACYSSELSKRPCTASSG--GS---IARGA 1526
Db 2006 LGQLPQGBR-----DSAVDLSLK--HYSISLGADRYLGGLQYQSFPDLRPT 2054
QY 1527 PVIVBELKPROSPLT--YEDHGAPFAGHLPRGSPVTNREPTPLQEGSLSSKXADRK 1584
Db 2055 DLHSHPLRLKRYSSVSNYSDH-----RYGRGDAV-----GRQASLQYSA----- 2097
QY 1585 LTSTPREIAK--SPHSTVPEHHHPISPYEHLKGVSGVDLYRSHIPLAPPTGISPRGI- 1641
Db 2098 --TTAREISRMCAALNSMDQYGRHGS-----GSGGPDLYQ-YQPGHGLSAPQGLA 2147
QY 1642 PLDAAAYULPRHLAPNPTVPHLYPRYLIRGYRPTALENRQTIINDVYITSGQHHNTAT 1701
Db 2148 PLRSG-----LIGNPTP-----EGPSPGMLAQYGPASQATAVRQLLPSTAT 2191
QY 1702 AMA-----QRADMLRGLSPRESSLALNTYAAGRGIIDLSQVPHL 1740
Db 2192 VRADGMISTINTPIATLPITTPQPAVLBPV--RGGMVRYPGVGAVPLTSLTRV 2249
QY 1741 PVLVPTPTGTATAMDRLAYIPTABQPFSSRRSSPLSPGPTHLTKEPTTSSSRERDR 1800
Db 2250 FMIPRVPLGAG-----LYRYPAPRPPIASSVPPAE-GPVYLGKAAAKAGAGAGRP 2301
QY 1801 DREDRDRERKESI-LTSTTVEHAPIV-----RPCTEBSGSGSGSGGCGSSRPA 1852
Db 2302 RPELPAGVAREEPSTTAPAVIKAPVAPARAPAPRPQKPRAGEAAGSGSVLSRPA 2361
QY 1853 SH----- 1854
Db 2362 SEKEASQEDRQKQEQQLQLEHREVELKLRQLQLELERERVELQRHEEQQLVQ 2421
QY 1855 -----SHAQHSPISTODALQORPSVLHNTMKGI-----TAVBPSK-- 1894
Db 2422 RELQELQTIKQHVLCQQQEEBROAQFALQREQLAQRLQLEQIQLOQLOLEBOKQRO 2481
QY 1895 -----PTVLRSTSTSPVPAAT-----PPATHC-----PLGGTLGCVPTLMEPV-- 1937
Db 2482 KAPPPATCEAPSKGP-PAATIELAQNGQVMPPLTHAFLIYAGTEGQGP--REPVLMR 2537
QY 1938 -LPEAPRVARPERPRADTGHAFLLAK-----PPASGLEPPASSPSKGS----- 1979
Db 2538 GLPESASDMSLQTEBQEWAGSGIKKHSWRLRACGSPSGPDSTVRIADSSVQDD 2597
QY 1980 ---EPRPLV-----PVUSGHATIAFTPAKMLAPHASDP 2011
Db 2598 EEEGRYLVTRRRRTRRSADCSVQTDDBDNADWEQPV-----RRRSRLSRHSDGSD 2650
QY 2012 PAPPASADPHREKTKQSKPFSIOELSLSLGHSYSREGVEVPSP----- 2058
Db 2651 SKHDHTASSSTTAATAPAMS--SVGIOTISDCSVQTEPEQLPRVSPAIHTIATDPVE 2708
QY 2059 -----VSSPSLTHDKGLPKMLE--ELDKSHLEGELR 2087
Db 2709 IVRYISAEKTRGEGSLACQTEPDQAGVAGPQLIGFTALISPLYRGIQIYTPALGRFE 2768
QY 2088 PKQCGPVLGGEAAHLR--HLRPLPESQPSPLQTPAKYKGNRVYVTLAQH----- 2138
Db 2769 KKKDPLEIGYQ-AHLPEPESLSQVSRQPKSPQVLYSP-----VSLPLPHRLLDTSF 2820
QY 2139 -ISEVITQDVTGRHHPQUS-----APLPALYSPGASCPVLDLRPPSDLYLPP-- 2188

```

```

Db 2821 ASSERLKAHVSPQKFIADSTLRQOTLRPPMKT-----LQRLSD---PKELSP 2867
QY 2189 --DHGAPARSPHS--EGKKSPEPNKTSVUCGEDGIE-----PVSPEGWEP--GH 2236
Db 2868 TAEESAKERFSLYOHQGLGS---QVSVL--PBNGLVRVKRTLPSPPEBAHLPLAQ 2921
QY 2237 SRSADV--PLLVRDGEQTEPSRMGSKSPGNTSQPPAPFSKLTESNAMYKSKQIINKL 2294
Db 2922 VPSQLYASLLQR-----GLAGPTVPATKASLLAEELDRDLALVEHESKAKKKQ 2971
QY 2295 NTHNRPENYINISOPGTEI-----FNNPAITGTGLMTYRSQAVGEHAS-- 2337
Db 2972 AEIDEEKEIDAKLYBELGITYOKESLAKORGRDVPPLRGIG-----EHRDYL 3021
QY 2338 -----TNMGLEAIRKALMGKYDOMESPPLSANAFNP--LNASASLPAAMPITLADGRS 2390
Db 3022 SDSELNQLRLQGGCTTPA--GQYVDYPAASAVPAPPPSGPTAFQOPRPPAPQYTAGS-- 3076
QY 2391 DHTLSPGGG---KAKVSGRPSPRSKAKSPAPGLAGDGP--PSVASVSEGDGCRNRT 2444
Db 3077 ---SGFTONGPFAHQAPTYTGESTYPAPTYPRCTGYPAEBGLPSQAFHPTGYAAPTP 3132
QY 2445 L-----TNRVWEDRPSSAG---STPPVNPILMRLOAGVWASBPPEGLPAGS 2488
Db 3133 MPTQSAFPFVQADSRRAHQKPRQTSIADLEQKVPYTVYVIGSPAVYMSAPPE--TGYS 3190
QY 2489 GPLAGPHHAWDEEPKPLCSQYETLSDS 2516
Db 3191 GPVAVSGSYEOKKAPENHPRGSDRSVSQS 3218

RESULT 11
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
R:label: S., Kolmerer, B.
Science 270, 293-296, 1995
A:title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: J38346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; NID:91017426; PIDN:CAA62189.1; PID:91017427
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 3.8%; Score 503.5; DB 2; Length 7962;
Best Local Similarity 19.0%; Pred. No. 8.2e-10;
Matches 534; Conservative 355; Mismatches 1053; Indels 865; Gaps 128;

QY 27 PVQIARTHTDVGLLEYOHHSRDVASHLSPGSIQPPQRRPSLSLSEFPQGNERS-----Q 80
Db 5498 PLOFTKIQNVASE--HQSATFCEVSPDAIYTWYKGPTELSTESQKYNFRNDGRCHYM 5555
QY 81 ELH-LRPESHSHYBELGKSEME-----FIESKRPTEL-----LPD----- 115
Db 5556 TIHWNTPDDEGVYVIRALEPRGARSTABLYLTTKGIELKLEKRPDIPDSVPIPTMIR 5615
QY 116 -----PLLRSP-----LLATGQPAQSED-----LTQRSLTGKLEPV 148
Db 5616 AVPEEILPVVAPVPVPLLPTEBEKKPPKRIEYTKKAVKADAKVAKPEKMPREBIV 5675
QY 149 SPSPRPHTPDELEVPRLSLKEELIQMMDRVDRITWVEQIISKL-KKKQOQLEEAAP 207
Db 5676 KKPPPTT-----LIPAK--APEIIVDVSSKAE-----EVKIMTTTKKEQVKEKAV-- 5720
QY 208 PEPEKVPSPPIESKHSVLQIYDENRKAEAHRILEGIPQ-----VELPLYNOPSD 262

```



```

Db      2352 LVKTTKSKPAEMTBRRLDKLKEMVPHSEYEAFAVPAHGEODES--GSPNNMFLQPLD 2410
Qy      943 SLTTPGPRRA-----NASPO-----KPLD---LQQLQORAAIPIQVT 979
Db      2411 QELTSSQPHGWNPHNPTGKIYLHYAEPPTGPFVEBPDLFLTKTKSKPVQSGPKELA 2470
Qy      980 K-----VHEPRREDA--PTKRAPAPAPP-----ONLQESDAPQOPGSS 1018
Db      2471 KSPKEMVQOTPEYKBAVLSCPEODESPSPNNMLKSLDQEVMSQPHSGVPRP--KT 2529
Qy      1019 PRGK--SRSPAP-----PADKEAPAAEQCLPDDPCWISGLPPVPRPREVIR---A 1065
Db      2530 P-GKIYLSIEBPBPFPVKEPTDLILVKTTKSKPAE--WT-----PRRIDKLKEM 2577
Qy      1066 SPHADPFAFSYAPRGHPLRPLGHLDTARVPLRPRPTISNPRPLISSAKHPSYLEQIGAI 1125
Db      2578 VPHSE-----VEBAVFP--AHGEQDESGSPNNMFLQ-----LDQELTSS 2617
Qy      1126 SQGMSVOLHVP-----YSEHAKAPVGPV-----TMGLPLPM--DPKCL 1161
Db      2618 SQPHWVPHNPTGKIYLHYAEPPTGPFVEBPDLFLTKTKSKPVQSGPTQMAKSPBEM 2677
Qy      1162 APFSGVQEOQLSPRQOAPRPSLGVPTAOEASVLGTAAGSVPGSITKGIPSTRVPSDS 1221
Db      2678 VLSLSPKNEIYFPAQKQODESPSPNNISLQSLDQETMSQPHGMIPIH-----PPN- 2729
Qy      1222 AITVAGSTTHG-----TRADVLYKGTITRIIGEDPSRLDRGREDLSLKG 1266
Db      2730 -----THGKIYLHYAEPPTGPFVEBPDLFLTKTKSKPVQSGPRQIDSKPKMTQ-- 2780
Qy      1267 HVIYEGKGAHVLSEYCGMSVTQCSKEDRSSGPHETAAPKRTYDMEGRGVRAIISAS 1326
Db      2781 -----SPYEESLPLPAHAEQGESRAPRPHSLQF-----LDQELTSS 2818
Qy      1337 IEGLMGRAIPPERHSRPHLKQEOHIRGSIITGIPRSYVAEOEDYLREAKILKREG--- 1382
Db      2819 -----HPHGMVPHNPTGPKIYLHYAEPPTGPFVEBPDLFLTKTKSKPVQSGTATK 2869
Qy      1383 TPRPPSRDLTEAKTQAL--GP-----LKLKPAHEGLVAIVYKGRSIH--E 1427
Db      2870 TDKSDEDEVSGQPEYKEAVLSGPGDQDBSPSPNNMSLQSLDQELTSSQPHGMIPIH 2929
Qy      1428 IPREELRHTPELPLAP-----RPLKEGSIQ-----GTPLKYDTG 1462
Db      2930 APDKIYLHYAEPPTGPFVEBPDLFLTKTKSKPL--QGTPTQMAKSPKEMVQOTPEYKAD 2988
Qy      1463 ASTTOSKGDVRSGLSGPQRTFPPVHPLDVMAADALERACYEESLKRPP--GTASSS 1518
Db      2989 LSAPEKNDK-----SPSSPNNMSLHPLD-----QELTSSQPHGMIPIH 3030
Qy      1519 GGSF-----ARGAPVIVPEL-----GKPRGSLTYEDHGAFFAGHLPRGSPVTRE 1564
Db      3031 HGKIYLAHAEPTGPFVEBPDLFLTKTKSKPVQSGPQIDK-----SKPEVFS 3079
Qy      1565 PTPRLQESLSSSKASQ-----DRKLTSTPREIAKSPHSTVEHHPH--PIS 1609
Db      3080 QSPSEESVLAQOAGEESRAPRPHMSLQILDQKLT-----LSSHPHGMIPIH 3133
Qy      1610 PYEHLKGVSG-----VDLY-----RSHTPLAFDPTSIIRG-----IPDA 1645
Db      3134 IYLAHYAEPPTGPFVEBPDLFLTKTKSKPVQSGPQVQSPKEMVQOTPEYEEVSLPQOA 3193
Qy      1646 AA--AYULPRHIANP-----TYPH--LYPP-----YLIRGVDTALLENRQIIL 1666
Db      3194 EGQESRAPRPHMSLQPLDQDILLSSHPHGMIPIHNTPTGKIYLHYAEPPTGPFVEBPDL 3253
Qy      1687 NDYITSQCMHNHTATAMAQADMLRGLSPRESSLALNAYAGPRGIIIDLSQVPHPLVLP 1746
Db      3254 FLRTTKSKPAQMTPTQMAKSPBEMVSLSPENKETY--PPAQKQOEEISIPPHMSLQPLD 3311
Qy      1747 TTPGPAITMDRLAVIPLTAQPPSS--RHSSPPLSGGP-----THLTKPTTSS 1793
Db      3312 QDLTPSSHPH--GMIPIHNTPTGKIYLHYAEP--PTGPFVEBPDLFLTKTKSKPVQSP 3367

```

```

Qy      1794 SERERDRREDRDREREKSLTSTTYVHNAPIMRPTGEOSSGSSGGGSSRRPAS 1853
Db      3368 RQIDSKPEVFTQSEYEESVYL-----PAQAGEGE----- 3398
Qy      1854 HSHAHQSPISPRIGD--ALQORPS--VLHNTGMKGI-----TAVEKSPETVRS 1900
Db      3399 -SRAPPHMSLQPLDQDILLSSHPHGMIPIHNTPTGKIYLHYAEPPTGPFVEBPDLFLKT 3457
Qy      1901 TS-----TSSP-----VRPA-----ATPP-----A 1916
Db      3458 TKSKEVHSGPRQIDKSHKEMFTQSEYEESVYLPAQAGEESRAPRPHMSLQPLDQDILL 3517
Qy      1917 THCPGL-----GTLGVPYPLMEVALLP-----KEAPRARERPRADTGH 1957
Db      3518 SH-PHGMIPHNPTGKIYLHYAEPPTGPFVEBPDLFLTKTKSKPVQSGPRQIDSKPE 3576
Qy      1958 AFLAKPRASGLEPASPSPK--GSEPRPLVPVSGHATARTPAKDLAPHNA--SPD-- 2010
Db      3577 MFTQSPYEESVLAQOADEESRAPRPHMSLQPLDQDILLSSH--HGMIPIHNTPTGKIY 3635
Qy      2011 ----PRAPRASADP--HREKTSKPFSLQELSLRSLGYHS-----SYSP 2050
Db      3636 LHYAEPPTGP--FVEBPDLFLTKTKSKP-----VHSGPRQIDSKPEVFTQSP 3682
Qy      2051 EGVEPVSVSSPSLTHDKGLPKH--LEELDK-----SHLEGL--RPKQRPVULGGBAA 2101
Db      3683 EYESVLAQOADEESRA--PRHMSLQPLDQDILLSSHPHGMIPIHNTPTGKIYL-- 3736
Qy      2102 HLPLRLPSESQPSRPLQTPAGVGHQAVVTLAQHISEYITODYTHNHPQOULSAPRA 2161
Db      3737 HYAEPPTGPFVEBPDLFLTKTKSKPVHSGPRQIDSKPEVFTQSP--PEYESVLA 3791
Qy      2162 PLYSPFGASCP-----VIDLRPPSLYLP 2189
Db      3792 QAEDQESRAPRPHMSLQPLDQDILLSSHPHGMIPIHNTPTGKIYLHYAEPPTGPFVEBP 3851
Qy      2190 -----HGAPARSP-----HSEGKRSPEPKTVS--LGGED-----GIEPV- 2225
Db      3852 LFLRTTKSKPVQMTPTQIDKREBMSQSPENEBADIPRHGDQDELINRPIISLQPLD 3911
Qy      2226 ----SPREG--MEPGRSRAVYPLLYRD--GEOTEPHMSKSPGNTSQPAFFSKL 2275
Db      3912 QELTSSQPHGWNPHNPTGPKIYLHYAEPPTGPFVEBP-----PDLFFLRT 3957
Qy      2276 TESN-----SAMYKSKQELINKKLTNNHNEPYN--ISQGTETFMNPAITGTGLMT 2326
Db      3958 TSKRPVQGTPTQIAKSPKEMVFO-----TPYEKAVLSAPGEDODETP----- 4000
Qy      2327 YRSQAVQEHASTNNGLEAIIRKALMGKYDQ--WEESP-----LSANAFNP----- 2370
Db      4001 -----SPNNISLQSLDQELTSSQPHGMIPIHNTPTGKIYLHYAEPPTGPFVEBP 4051
Qy      2371 --LNASASLPAMPITAAADGRSDHTLTPSGGGAAYVSGRPSRKAAPAGLASGDRP 2427
Db      4052 LFLTKTKSKPVQGTIVQMD-----KSPKQGL----- 4079
Qy      2428 PVSASVHSEGCNRTPLTNKWE--DRPSAGSTPF-----PYNPLMLRQOAVMASPP 2480
Db      4080 ----PAHGAQADESRSPNNMSLQPLDQELTSSQPHGMIPIHNTPTGKIYLH--YAEP 4132
Qy      2481 ----PPGL-----PAGSGPLAGPHNMADEPRKLLSQVETLSDSE 2517
Db      4133 TGPVEBPDLFLTKTKSKPLAGITPTQMAKSPKEMFS--QTLBNKE 4176

```

RESULT 9

S36152
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
 C:Species: Homo sapiens (man)
 C:Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
 C:Accession: S36152
 R:itis, F.J.M.; Bougueteloret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Jurka

Query Match 3.9%; Score 520; DB 2; Length 2187;
 Best Local Similarity 19.7%; Pred. No. 5.4e-11;
 Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;

QY 738 GPATNNSSDTEISIPHTAA-----KDTGONGKPPATLGADGPP--GGPPT 784
 DB 3 GEATITVATEBELPOQAFETAVLPMSSALKVAAGVGPPTPPSSILGQCSQSVTAHOPS 62
 QY 785 P-PRRTSAPRIE-----PTPASEA-----TGARTP-----PP----- 810
 DB 63 PLPSSVSTPFEVPPAQTITTAETLPSGTAPPTPTPLPHLIGPISPAALALASPMIGLA 122
 QY 811 -----APPSGAPPPVVPKEKEKEETAAAPVE----- 838
 DB 123 QKGAUSSAPLSLVALPHSVQSSVCPPHPLTSPPSAAGALGLMTLSIPLEFKTSTS 182
 QY 839 -----EGEEOKPP-----AAELAVDTGAEEPPVXSE 865
 DB 183 QVPSQGITNLKGTACPPDVVRAFPSSHLENPASVQPGIMSCPQTLSMTSPVKGVPISSA 242
 QY 866 CTEABEOPKAGKDAEAAEATAGALKAE-----KKEGGSGRAT 904
 DB 243 LTQSLSLNLKGPVPPRANTAPASIPLA PSTSLCCH.PLHSSVDSPTIQPGQSGLAIV 302
 QY 905 T-----AKSGAPODSDSSATCSADEVDEAGCDK- 934
 DB 303 SNPTVGHSGIAASCPRECVVPALPSRLAVDSGAAPSDDKSSAVINELCSPPGSSNV 362
 QY 935 -NRLLSPPSLITPTGDPRAANASFOKLDKOLKORAAIPPIQVTKVHEPP----- 985
 DB 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASQCTGKEKIFVSCIGATH 409
 QY 986 -----REDAPTPAPAPPPPPQNLQPSDAPQPGSPRKSRSRPAADKEAFAAEAKX 1041
 DB 410 HALNDPSALISVAPATHVPPPT-----SGGLVSKDPAASVPTSLVPPAAKQ 455
 QY 1042 LPGDPCWCTGLPF-PVPPREVIXASPHAPDSAFSVAAPGHPLPLG-----HDTARFVL 1096
 DB 456 FPAPASATLGVVPSFLPATEGLK---NLPI SALVNGAVPSPAQAGLPTRKDTTLQOLA 512
 QY 1097 P-----RPPTISNPPLISSAKHPSVLEROI GALSQMSVOL 1133
 DB 513 PIALKESPSOSASLEVLSEDTVTYKTTGGPAPVVRPA-----IAGVATTTSLARA 563
 QY 1134 HVPYSE-HAKAPVGVNTWGLPL-----PMDPKLIAPSGVQKQOLSPRGAGPPE 1182
 DB 564 DSPRAVTAADSCVSNTVTSQPLKBSVTPPAMAFTAKTAPST---TSPVLPLASEGCPV 620
 QY 1183 SLGVPTQAEASVLRGTALGVPPGSGITKQIP-----STRVPSDAITTYRSGI 1229
 DB 621 ASSMALSPONASVSFTALALSP--EIPKSVPPDPPLAEIFSNARKVDVAVSHMSSGSS 678
 QY 1230 THGTP-ADVLYKGTITRIIGEDSPRLDRGRED-----LPKHVIT 1269
 DB 679 RQGHDPDASVYAKGVVCL-----ADSSLDTSVASASGALSASGASPLYLEVSLPEAGLA 734
 QY 1270 YEGKGHTLYEYEGGMSVTQCKEDG--RSSGSPPHETA----- 1306
 DB 735 VQGRKGL-----NKLSTPPSSKGAAPVSTGAPPKAPIVPTSSSISQVPAELIPS 790
 QY 1307 PKRTYDMMEGVRAISSASILEGLM-----GRAIPPERHSPHMLKEOHNT--- 1351
 DB 791 POKTPEVTAASRLISAVQSPKVPDIPMSDVTPTSPKKTSAVAPKOTSATLSKSPAVATSL 850
 QY 1352 -----RGSITQGIIPR----- 1361
 DB 851 SPKAPVAPASNEATIVPTPIPTSLKNALAAATPKETLATISIPKVTSPSPOKTPKSVSLKG 910
 QY 1362 -----SYEAOED-----YLREAKLTKRE-----GTPPP--PP 1388
 DB 911 APAMTSKATETIAASKDVSPSQPFKEVPLQHVPTTSPKPSVSDTLGALTSPPPKGPP 970
 QY 1389 PSRDLTEAVYKTQALGPLKLK-----PAHEGLVATVKEAGRSIHIEIPR-----EELRHTEPL 1439

DB 971 ATLAEPTTPYKPSKPPKASKKTTPATPSPEGTAVALEIPCSKAKPXTAAKRESSATSSS 1030
 QY 1440 PLAPPLKEGSITQ-----CTPLKDTGASTTSSKKHNDVRLIGSPERT----- 1483
 DB 1031 KRAPKTAASKEIPSKGVTAVPLEISLPLKETSKSATPDEKS-----ASSPKRSPKTAGP 1084
 QY 1484 --FPP-----VHPDVMAADABALERACEYESI-----KSRPGTA-----SSSGGSJAR--- 1524
 DB 1085 KETPPGVTAVPPELSLPPKKTPOMATPNEISLAASSQKRSKTSVPKETPPROGVTAEMLE 1144
 QY 1525 --GAPVIVPELQKRPQSPLYTEHDGAPPAHLPRGSPVTMBEPPTPRLQEGSLSSSKASQD 1582
 DB 1145 IPSAPQKAPKTAVPKQIP-TPED-----AVTLIAGSLSPKK-----ASKTAAP 1187
 QY 1583 RKLSTPR-----EIAKSPHSI-----VPHHHPHPIPTYHLRGVSGV 1621
 DB 1188 KEAPATPSGVIAVANGSELSPSPKKTSKTAPKENSATLPPKRSKPTAAPKETPATSSBGV 1247
 QY 1622 DLYRSHIPLADPPTSIPRGIPLDAAAAYYLPRHLAPNPTYPHLYIRGVPTPALEN 1681
 DB 1248 TAVSELSPS-PPTPASKGVV--TLTPKGAENALAESPASP-----KVPKTAABEE 1297
 QY 1682 RQTIINDYITQOMHNTATAMAQRADMLRGLSPRESSIALNTVAAGPRGIIDLQVPHLP 1741
 DB 1298 TST-----TSPQKIPKVAQ--PKRAS----- 1317
 QY 1742 VLVPPTPGTPTAMDRLAYLT-----AAQPSRSRHSBPLSGPHTHLTKPTTT 1791
 DB 1318 -ATPPSKTPTTAVPKETASPSBGVTAVPLEIPSPKRAPPTAAP-----KETPA 1366
 QY 1792 SSSERERDRDRDRDREREKSIILSTTVVHAHIMRGTGEOSSGSSGSGGSSSRP 1851
 DB 1367 PSPEG-----ATTAPVQIIPSPKSGSKA-----GSKETP 1336
 QY 1852 ASHS-----HAQHSPISPRTQDALQORPSVLAHNTGMKGIITVAEPPKTYLRSTSTSP 1906
 DB 1397 TTPSBEGVTAAPLEIPSSKTSKMASPKETLVTPSSKGLSQVGPKEKTSLEGATVAVLE 1456
 QY 1907 VRPAATPPPAT-----HCPGLGTLGQVPTLMEPVLPREARVARVPERPRADTGHAPLAK 1962
 DB 1457 IPESHKAPKTVDPKQVPLTPSPKDAPTTLAESSPSPKAKTAAPESEKVTY-----V 1510
 QY 1963 PPASGLEPASSPSKGESEPRPLVPVSGHATITAPKAKNLAPHNASPPRPAAP--SASD 2020
 DB 1511 PP-----EKATPQKASGTTASKVPV-----PAETQVANSSRPTPTPAVPPVKN 1556
 QY 2021 PHREKTSKPSIQLLELRSLGYHSSYSPEGVPSVSSPSLTHDKGLPKHLELDKS 2080
 DB 1557 PSHKKTSTKTELKE-----APATLPP-SPTKSPKIPSSKKAER----- 1594
 QY 2081 HLBELERKQGPVYKLGGEAHLPHLRPLPS-QPSSSPILQTPGVYKHGRVVTLAQHI 2139
 DB 1595 ---TSAVK-----EPASPSIKPVTTSLAQTAAPSLQKAPS----- 1627
 QY 2140 SEVITQDYTRHHPOLASPLPAPLYS--PPGA-----SCP 2172
 DB 1628 -----TTIPKKNLAAPALVPSKSPAPAPASASLSAPATAQCTAPKEATTIISCK 1679
 QY 2173 VLDIRRPPSDLYLRPPHGAARGSPHS-----EGGKRS-----PEPNK 2211
 DB 1680 KAAATEPIENTSTAPLEGAPKETSETSVSKVLMSPPKXASSKRASLTLPATLLPLKE 1739
 QY 2212 TSVL-----GGGDDGIEPVSP--PBGMTGHRSAVYPLLYVDGQCTEPRSGMSGSPG 2263
 DB 1740 ASVLSPTATSSGKOSHSPVSDACSTGTITTPQASEKL-----PSKKG----- 1781
 QY 2264 NTSQPPAFPSKLT--EENSAMVSKQOEINKKNTAHNRNEPEVNIQOPTEIFMPPAITG 2321
 DB 1782 -----PTAFTENLAPAPESALATAPIQKSPGANSNSASBPCKDPSKDTKLPS--- 1844
 QY 2322 TGLMTYRSQAVOEHASITNMGLEALIRKALMGKTDOMEESPLLSANAFNPUNA----- 2373

```

Db      653 QQHQQSGAPPTVPSPPOQVTL--GAVPAPQAP-----PPPKALY----- 692
Qy      1043 PGDPPTCTSGIPFVVPREVVKASPHADBSAFYAPRGHPLPGGLDHTARPVLPRTPTI 1102
Db      693 PG-----ALGRPPMPMNPF-----DPRMMIIPVVD-----RLLOGRPBLE 730
Qy      1103 SNRPPLISAHAPSVLERQIGATISQMSVOLHVPYSEHAKAPVGPVTMGLPLPMDPKKA 1162
Db      731 FYRPGV-----HPSGLVPRKSDSLGLSS-----PFDHAPAMML--ERGTP--PVD--KLA 778
Qy      1163 ----PFGSVKQEQ-----LSPRQAGRPESLGVPTAQEASVLKGTALGSVPGSITKGIPST 1215
Db      779 WVGQVFATAPAEPRPLSLRQAADDDKGM-----RSETPVP-- 818
Qy      1216 RVPSDAITTRGSIITGTPADVLTKGTTTRI--IGEDSPSLDRGREDSLPGHVIYEGKK 1274
Db      819 --PPVYLASYGFENAGAP--PISRPLEPGRPL-----FWPG----- 857
Qy      1275 GHVLVYEGGMSVTCSEKEDGSSGPRHETAPRKTYDMEGRGVRAISSASIEGLMRA 1334
Db      858 ----SDEVAKITPPPKPEPKKEITQLTGPEAGKULPASR-----SGAG 898
Qy      1335 IPRERHSPHLKEQHNRIGSITQGIPIRSYVAQEDYLAREAKLKRBTGP-----P 1386
Db      899 PPPRRRSRTETRKPRRSGSRGIP-----EPPGAPRRRAGPIKKP 941
Qy      1387 PPPSRDLTEAVKTOALGPLKAPHEGLVATVKEAGRSIHEIPREBELRHPLEPLAPPL 1446
Db      942 PPPPK-----VEELPPKPLECGDETEKPPRP- 967
Qy      1447 KESITQGTPLKYDTGASTGSKKHVRSLLSGPRTPP--VHPLDVMADALER-- 1501
Db      968 ----DLKLTIKG-----LGGPKET--PENGUISPAPRLRDIYSYERVG 1005
Qy      1502 --ACYEESL-----KSRPTASS--GGSITARG--APVIYE 1532
Db      1006 PTCRGRGRGRTYFARGRFRGTGGRGQGANLAVTESFEMMGVEVGQDDQTLLLE 1065
Qy      1533 LGKP-----ROSPLYTEDHGAPRAGHLPRGSPV-----TMRPPLRLOEGSL 1575
Db      1066 AAMPARHAGAVQSMKSP-----SGAGSGAQOKAPRPMRVIMLQTRLLPFRREHSPRS 1120
Qy      1576 SSKASQBRKLTSTREIAKSPHSYVBEHHPHPISEYHLLRGVSDVLYRSHPL--APDP 1634
Db      1121 S-----RSPPTNSPTLHR--APARTCPGVG-----ESSLEPGALSP 1155
Qy      1635 TSIPRGIPLDAAAYVLP--RHLPV--PTYPHLYP-----YLIRGYPTALLENROT 1684
Db      1156 GPRRREARPOVCPGMSPPAKSLAPKXPTGP--LPSPKEPLKXLIPLPLSPVARCGSNG 1213
Qy      1685 IINDYITSQMHNTATMAQPADMLRGLSPRESSLAL--NYAAGPRGIIIDLSOVPHLP 1741
Db      1214 GSNVGMEDGERRRRRHGRAGQODK-----PRFRRLKQERENAAAGSEG----- 1258
Qy      1742 VLVPPTPTPTA--MDRLAYLPAPQPFSSRHSSSLSGGPHHLKPTTSS----- 1793
Db      1259 --KPSLTPASAPQPEBALTTVVAPAPRAAKSPDLSNONSQANEWEITESSEDF 1315
Qy      1794 --SERERDRRE-----RDREERKESILTSTTT 1820
Db      1316 TSERRGDEARPVVLTPTKAVGTGCGGAGVPGISMSRDDLQGRANDLKRSSSGRP 1375
Qy      1821 VEHAPIWRPTEQSSGSSGSGGGG--SSSPASHSHAHQ 1859
Db      1376 GMEQRNRRPGCGKXKSSGSSGSGGGGPGGRTGGRGDKRKSWSPPKXRSRPPREHPPGL 1435
Qy      1860 HSPISPRQDLOQRPSVLTHTGMKGITITAVEPSKPTVLASSTSSPVPRAPATPPATNC 1919
Db      1436 PLPPPPSSSAVFRIDVYH-----SNPAGIQALQLSLRQGSVTAPGCH- 1481
Qy      1920 PLGGTLGVVPTLMEPVLLPKAPARVAPREPRADTGHAFKAPAR-----SGLEPA 1972
Db      1482 -----PR-HKGPQAPQGS--PRPRTREPORVNSGL--- 1512

```

```

Qy      1973 SSPSKGSEPRPLVPVSG-----HATARTPAK-----NLAPHHAS----- 2008
Db      1513 SSDHFEPRPMPRKGCTGRDAGVSPPPKRRERPRKRELLQESLPPSHSSGLGS 1572
Qy      2009 -PPRPPAPASAD-----PH-----REKTQSPFSIQ-----ELE 2037
Db      1573 KPGEPPQASNRDTGTALPRLIMNRLHTATSRKSYRPTSMPEMMEPLSPFEDVAGTEMS 1632
Qy      2038 LRLSGHGSSTSPGCVPEVPSVSSPSLTHDKGL-----PKH-----LELDKSHLEG 2084
Db      1633 QSDSGVLDISGDQVSSSGSCQSRSSP-----DGLKGAEGCPKPGGSSPLNAPCEGPPG 1688
Qy      2085 ELRKPQCPVTKGEAHNLPHLRLPLRPSQSSSPPLQTAPGVGHORVTLAOHISVIT 2144
Db      1689 SEPRRRPPAPHODKRELREPOLP-----DGPITGR-----SOR-----T 1726
Qy      2145 QDYTRHHPQOLSAPLPAPLYSPGASCPLDLR-----PPS-DL 2183
Db      1727 DRGTEPRPIPS--HRPGPPVQF--GTSKDSDLRLVGDLSLKAKEKLTASVTEAIPVSRDW 1784
Qy      2184 YLPPDHGAPARQSPHS--EGGRKSPEN-----KTSVLGGEGDT- 2222
Db      1785 ELAP-----SAAASAPQSKNLDSGHCVPEPSSGQRLYPEVFGSAGPSSQISGSGSHGLS 1841
Qy      2223 -----EPVSPREGMTEPGHSRAV--YPLLYRQGEQTEPRMGSKS 2261
Db      1842 ITSQKMLRPGTSLPHYSQPLYPGPAPPSALLSGVALKGQFLDPTMQATELGKL 1901
Qy      2262 PGNTQPPAPF-----SKLTESNAMYKSKQKQINKLNTNRNEPXYNISQPGTEI 2313
Db      1902 AGGVLYPPPSFLVSPACPSPLPDTSLQVRQ-----DLRSP--SDF 1941
Qy      2314 FNPFAITG--TGLMTYSQAVQEHASTNMGLEAITRKALMGKXDOMEESPPLSANAANPL 2371
Db      1942 YSTPLQGGGSGFLPSGAPAQO-----MLPMDVSQLPVNFGSLPAP--PAPPL 1992
Qy      2372 NASASLPAAMPITADRSHTLSLPGGGKAKXVSGRPSRKAASPAAGLASGDRPVS 2431
Db      1993 SLPLVGPALQPSLA-----VRPPAPATRVLPSP-----RPPPAS 2029
Qy      2432 SVHSE-----GDCNRTPLTNRWEDRPSSAGSTPPYNPL--M 2469
Db      2030 LGRALHLPVELKPRQDQKLSNNGRSGSRTPTGFSFGSLNRLKATSTYSGVVRTQ 2089
Qy      2470 RLQGVNASP-----PPGLPAGSGPLAGPHHAWDEBP 2502
Db      2090 RVDLYQOASPPDALRWIPKEMERTGPPRGGPSRR--AEEP 2128

```

RESULT 7

```

T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yoclov, W.V.; St-Arnaud, R.
Gene Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: EMBL:U48363; NID:G1666680; PID:G1666689; PIDN:AA18732.1
C:Genetics:
A:Gene: Nac
A:Map position: 10
A:Introns: 24/1; 1966/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ.
C:Keywords: alternative splicing; DNA binding; transcription factor

```


Db DEAEKSEKESRESVAEQP-----LVKSEVRPASPVAESVVDDEAEKSEKESPLMSKASR 3149

QY 1776 PLSPGPGPHLTKPTTSSSERRDRDRDREREKSLTSTTVEHAPIWRDTEQSS 1835

Db 3150 PASVAG-----SVDEAEKSEKESR-RESVAEKSPLPSKEASRPASVAESVKDEAD 3199

QY 1836 GSSGGSGGGGSSSRPASHAHQHSPISPTDOLQ--QRPVLAHTGMKGITTAVERP 1893

Db 3200 KSKESRRESGAEKSPPLASKASRPASVAESIODEAEKSEKESRRESVAEKSPLPSKAS 3259

QY 1894 KPTVL-----RSTSSPVPRPAATFPATHCPLGLTDGYPTLMEVLLPKKEAP 1944

Db 3260 RPTVASVKDEAEKSEKESRDSVAEKS-----LASKASR 3297

QY 1945 VARDEPRADTGAFLAKPARSGLEPASD---SKGSEPRPLVPVSGHATTARTPAKN 2001

Db 3298 PASVAESVQD--EAEKSEKESRRESVAEKSPLAYAKASRPASVAESIODEAEKSEKESRR 3355

QY 2002 LAPHHASPDPAPPASASDPHREKTQKPFSTQIEILRSLGTHGSSYSPEGEVFPVS 2061

Db 3356 -----ESVAEKSPLASKASRPTSVAEVKDEAEKSEKESRDSVAEKSPLAS 3403

QY 2062 PSLTHDKGLPRHL--EELDKSHLEG--ELRPQKQPPVVLGDEAHLPHL-----R 2107

Db 3404 KEASRPASVAESVQDEAEKSEKESRRESVAEKSPLASKASRPASVAESVKDEAEKSE 3463

QY 2108 PLPESOPSSSPLLQAPGVKGHQRVVTLAHOISEVT--TODYTHHPQOLSAPLPAPLYS 2165

Db 3464 SRRESVAEKSPLAS-----KEASRPASVAESVKDEAEKSEKESRRESVAEKSPLPSK--- 3515

QY 2166 PPGASCPVLDIRRPPSDLYLPPPDHGAAPAGSPHSEGGKSPPEPNKTSVLGGEDGIEPV 2225

Db 3516 -----EASRPFS-----VAESVKDEAEKSEKESRRESVAEKSLSAKXA 3554

QY 2226 SPPEGMPTPGHSKSAVLYLIRDGQTPERSMGSKSPENTQOPAPFFKLTJENSAMWKS 2285

Db 3555 SRPASVAE-----SVKDEAEKSEKESRRESVAEKS-----LASKASRPASVAES 3600

QY 2286 KKOELINKLANTHNRNE-----PEYNISOPCTEIFNMPALTGTGLMYRQAOVEHAST 2338

Db 3601 VKDEAEKSEKESRRESVAEKSPLPSEKASRP-----TSVAESVKDEADK 3644

QY 2339 MWGLEAIRKALKMGKYDOWEESPPLSANAFNPLNASASLPAMPITTAADGRSDHTLSPG 2398

Db 3645 SK-----EESRRRESGAEKSP-----ASMBASRPFSVAESVKDET----- 3679

QY 2399 GGGKAKVSGRPSSRAKSPAPAGLASGDRPPSVS-SVHSEGDGCR-----RTPLTN 2447

Db 3680 --EKSKESRRESVTEKSPPLPS-KEASRPSTVAESVKDEAEKSEKESRRESVAEKSPLAS 3736

QY 2448 RWMEDRPSSAGST 2460

Db 3737 KE-SSRPASVAES 3748

RESULT 5

TS1023

hypoetical protein B7F21.40 [imported] - Neurospora crassa

C/Species: Neurospora crassa

C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C/Accession: TS1023

R/Schulze, U.; Aign, V.; Hohlsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura

submitted to the Protein Sequence Database, July 2000

A/Reference number: Z25286

A/Accession: TS1023

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-2649 <SCH>

A/Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40

A/Experimental source: BAC clone B7F21, strain OK74A

C/Genetics:

A/Gene: NCSP:B7F21.40

A/Map position: 6

Query Match	4.2%	Score 552;	DB 2;	Length 2649;
Best Local Similarity	19.3%;	Pred. No. 4.9e-12;		
Matches	406;	Conservative 268;	Mismatches 771;	Indels 658; Gaps 92;
QY 60	QPCRRLSLSEPRG--NERSQILHRPESHVLPRLGSEM--FPIESGRPRLLELPD	115		
DB 848	KPVKPKAKDELPPRVSSEBPVAAKPEEPKPVSEPKKEKAPELLPELLEPEEPIQR	907		
QY 116	PLRPPSP-----LLAGPAGSEDLTKRSLSTGKTLPPVSPSPPHDPE	159		
DB 908	PQRPQRPQRPQRPQAOAOAPRQPLLTAAPKSVTPPERKEAELVQDAELRQOQAP-TERQ	966		
QY 160	LELVP-----PRLSKEL-----IQNDRVREITWYEQISKL	193		
DB 967	LLAIPVTEENETSFLRREPQPKHEMDTGSGSLPLPVEEHKLVDDQDVMVD-----	1020		
QY 194	KKKQOOLEE---AAKPREPKVPSPPIESKHSLVQIYDERKKAELAAH-----	242		
DB 1021	-----VDDEGRDALQF--PCKPRLRDVSDGSESVQFPAPFQOPTSDGSSSRMDEDS	1072		
QY 243	-----RILEGRLPOVELPLVNGPSPDTRQHEMIKINQAMRKK--LILY	283		
DB 1073	EERTIEDDASVSGSVENHVELSTPTTDELPRFN---VKPMHSGSKRYKLSGQPGCGAF	1128		
QY 284	F-----KRNHAKRMQKQKFCQRYDOLMEALEKKVERIENNPRRATESKVRE	331		
DB 1129	FMGNIRDAEETAEGHEA---KQVRYTNDAVLRF-----TLSDDP--TAVKSR---	1173		
QY 332	YEEKQPEPIRKQRLQERPMQF---RVGORGSGLSMSAKRSHEVSELI-DGLSE-QENLE	386		
DB 1174	---NQFSQSDKDXGSGKGSHGSDNHGEGGRRRTTSRFTSLDEYALQESIREAQEKKE	1230		
QY 387	KQMR-----QLAIVPMLY-DADQQRKFTFNNMGVADPMKVYKQDRVMMW----	432		
DB 1231	KEEPRAQREKTYTDEAVIPEMITWDEEDHOLFDTGSL--PLY---KLVTWQAVP	1283		
QY 433	-----SEQEKTFREKFWQHPKNGLIASFLERTVABCVLYLYLTKNENYKSLVRSY	487		
DB 1284	YHVNFTAEAAKFEKAYIETPKQMGKLAHLEPNDDHSVIOYAAKKRELNLKERLKKQF	1343		
QY 488	RRRGSGQOQO-----QOQOQOQOQOQPMRSPSOEKEDEKEKEAKEKEEKREPV	538		
DB 1344	RRRKGGRKQKYNALVSELGNPENETTEDNGENGENSGRRQOPRAAAPSFGHEATPNA	1403		
QY 539	END-----KEDLKEKTDPTDGEDNDEKENAASKRKTKANQGRKKGI-----	582		
DB 1404	DSDGATPSATGRRRATGTTTPPKDSCAE--KEGVKGGRRARQPPADKEPVKPAQ	1460		
QY 583	-----TRSMANANSEALITPOOAE-----	603		
DB 1461	AIAPTPPAASGKTNARSKNSTVQNPENMPPTPIALGARVQNPMDVPDPSGMQPLA	1520		
QY 604	-----IASMELNE-----	611		
DB 1521	PAQOTPLASPERAPPTLATTISSEMAPPSLRPEPPAPASLPTFIEIGSSGPERIRTPQ	1580		
QY 612	--SSWTBEBEETAKKGLLEHGRMSALAAKAVGSKTYSQCNFAYFNKK--RQNLDELQ	667		
DB 1581	QASSWSVSESNDEPSSLRSFGTWSALANMGTKTQVNVNYYLRLQKGGKQMEQIAT	1640		
QY 668	QHLKMEKERNAARRKKKAPAAASEEAAFPVVEDEMEASGVGNBEHEWEEAEALHAS	727		
DB 1641	EADLK--KQGERRRPPTPSAGR-----KRDVSSSGHRLAAAEAE-----	1684		
QY 728	GNFVPRGECGPATVNNNS---SPTESIPSPHTEAKD--TGQNGKPPATLGAD--GPPPG	781		
DB 1685	-PQAKSEAAEPANOPFGRFQVTPLQASPVLAHTLVQSSPTWMPALPLPAPVAAASAPAS	1743		
QY 782	PPTPRRRTSRAPLETPPSAEATGATPPRAPPSASAPPVVPKKEKEETLAAAPVBEGE	841		
DB 1744	PQTPBR-----FGAPVSQJMSVPVHPL---RQPAATFVTERGSEPI---PV----	1784		

QY 2085 ELRPKQPGVKLGGEAHLPHLR-----PLPESQSSPILQOT--ARGVKHQRVVTLAQ 2137
 Db 2021 ESTROYEGP-----LHHYRSQGSFSPQOQPLPFPSSQSGMGQVPRTHLLITLAD 2071
 QY 2138 HISEVITQDTRRH-POOLAPL-----PAPLYSPGASCPLDLRRPPSLYLPPPD--- 2189
 Db 2072 HICQIITQDPAKNOVPSQASTSTQTSFALSSTP-----VRTTSSRYEPESQOT 2123
 QY 2190 --HGAP-RGSPH-----SEGGKRSPPNKTSTVLGGEDGIEPVSPREGMTEPG-HSRSA 2240
 Db 2124 VLHPRFGRVSPENLVKDSRGRGKSPERSHI---PSEPYEPISPQG---PAVHEKOD 2177
 QY 2241 VYELVLRDGEQTEPSRMGSGSPGNTSQPPAFSGTLTSSNAMYKSKQOELINKLNTNRM 2300
 Db 2178 SMULLSGRGVDPABEQRSDSRSPGISYLPSPFTYL-ESTSPMYKSKQOELFRKLNSGGG 2236
 QY 2301 EPEYNISQPGTEIFNNPAITGTGLMTYRQAVOEHAHTNMGLEAIIRKALMGKYDQWE- 2359
 Db 2237 DSDMAAQPSTETEFNLPAYVTTSGAVSRSHSPADPAS-NIGLEBDIIRKALMGSTDDKVED 2295
 QY 2360 ----SPPLSANAFNPLNASASLPPAMPITPAADGRSDHTLTSPGGG-GKAKVSGRPSRK 2413
 Db 2296 HGVVMSHPV---GIMPGSASTSV-----VTSSKARDEGEPSPHAGVCKEKLINKNSRK 2347
 QY 2414 AKSPAPGLA--SGDRPVSVSYSSEGCNRRTPLTNRVWEDRPSASSTPPVNPILML 2471
 Db 2348 SKSPITPQSYLGTERTSPSSVSHSEGDYHROT--GWAMEDRPSSTSTPPVNPILTIM 2405
 QY 2472 QAGVMASPPPGPLPAGSGPL--AGPH---HAMDEPKPLCSQYETLSDSE 2517
 Db 2406 ----LSSTPPTQIACAPSAITQAPRHOQNRKWEREPALPSAQYETLSDSD 2452

RESULT 3

T46489
 hypothetical protein DKFZp434M075.1 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C/Accession: T46489
 R/Diesterioeff, A.; Lauber, J.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A/Reference number: 223035
 A/Accession: T46489
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1047 <AAA>
 A/Cross-reference: EMBL:AL137641
 A/Experimental source: adult testis; clone DKFZp434M075
 C/Genetic:
 A/Note: DKFZp434M075.1

Query Match 9.6%; Score 1272; DB 2; Length 1047;
 Best Local Similarity 35.0%; Pred. No. 8.9e-38;
 Matches 401; Conservative 149; Mismatches 333; Indels 202; Gaps 54;
 QY 1470 KHDVRSILGSGRTTPPVNPLDVMA--ARALERACYE-----ESLKSHPGTASSGGGIA 1523
 Db 7 KHNKXSLITGPKSLSRGMPLEIVENIKVVERGKYEDVAKGTFVRSHTSVSSGSPVL 66
 QY 1524 RGAPVIVBELGKRPQSPFLTYEDHGA-----PRAGHLRGRGSPVTMRPEPTRLQEGSLSSK 1578
 Db 67 RST---LHEAPKAOISFGIYDITSAKRTPVSYQNTMSRGSFW-----MRTSDVTISSNK 118
 QY 1579 A-SQDRKLSTPRP-----IAKSPSTVPEHNPRIPEYENHLRGSVDLYRSHIPLAFD 1633
 Db 119 STNHERSTLTPTQRESIPAKSVGVGDVVSH--SPRDPHNRSTAGEVYRSHLPLHLD 176
 QY 1634 PTSTPRGRLPD-AAAAYLRLPHLARNPTPYRLPYRLKRGYRDLALDEN-RQTIINDYIT 1651
 Db 177 P-AMFRAHLDPPAAAYLFORQLSPTRYPSQYDLY-----AMENRQTIINDYIT 226
 QY 1692 SQQVHNHTATAMQRAVMLRGLSPRESSIALNVAAGRGITDLSQVPLPLVLPVPTGTP 1751

Db 227 SQQVQVNL-----RPDVARGLSPREQPLGLPYPA--TRGIIDLTNNPP--TILVHPGCTS 278
 QY 1752 ATAMDRIAYLTPAQPPSSR-HSSPLSPGCPHLLTKPTTSSSENRDRDRDRDR 1810
 Db 279 TPMDRITTYLPGQITPPRPYNASMPCHPTNL---AAAAGERRERERERERER 335
 QY 1811 EKSLITSTTVHAPLIMRPGTEQSSSGSSGGGGSSSRPAHSHAHQNSPISPTODA 1870
 Db 336 ---IAASDL---YLPGSEQ-----PGRPGSHGVRSFSP-SVKTOET 373
 QY 1871 -LQORSVLNNTGMKGIITVAEPSSKPTVLSTSSSPV---PPATFPPTACPLGTL 1925
 Db 374 MLOORSVFGTGITGVITPTLDPADLRIMPLRAGGSSISQGLPASYNPAADA-LAALV 432
 QY 1926 DGYYPTLMEVLLPK---EAPVANPERPAD-----TGH 1957
 Db 433 DAASAPQMDVSTKSKSKEAALLENLRSAVSEQQOLEQKTEVEKRSVQCLYSS 492
 QY 1958 AFLAKPARAGLBPAS---PSKSEPRPLVPVSGHATIAFTPAKN--LAPNHASDP 2012
 Db 493 AFPSGKP---QPHSVVYSKQKGP--PKRSYEEELRTRGKTTTAAANFIDVIT 545
 QY 2013 APPASADPRREKTQSKPFSIOELYSLSGYHSSSVSPGCVBPVSPSLTHDKGLPK 2072
 Db 546 RQIASDMDAREKSSQSDS-----SSLSSHRYETPSDALIVISPAASPAPQEKLOTY 599
 QY 2073 HLEBLDKSHLEGELRPKQPGPVKLGEAHLPHLRPLPESQSSPPLQTPG----- 2125
 Db 600 QPEVVKANQAMENPTQYEGP-----LHHYRP---QOESPSPQOQLPSSQABGNG 647
 QY 2126 -YKHQQRVTLAHIHISVITQDTRRH-----HQQSLAPLAPLYSPGA--SCPVLDLR 2177
 Db 648 QVPRTHRLITLADHICQIITQDPAKNOVSSQTPQO--PPTSTFQNSPLSVSTPV---R 701
 QY 2178 RPSDLYLPPPD-----HGAP-ARGSPH-----SEGGKRSPPNKTSTVLGGEDGIEPV 2226
 Db 702 TKTNSRYSPEQSQAVSHQRPGRVSPENLVDSRGRSPKSPERSHV---SSEPYEPI 758
 QY 2227 PPEGMTPEGRSAVPLVLRDGEQTEPSRMGSKSPGNTSQPPAFSEKLTESNAMYKSK 2286
 Db 759 PPQ--VPVVEKQDLSLLSQRGAPAEQRNDARSFPSISYLPSPFTYL-ENTSPMYKSK 815
 QY 2287 KOELINKLNTNHNNEPPEYNISQPGTEIFNNPAITGTGLMTYRQAVOEHAHTNMGLEAI 2346
 Db 816 KOELFRKLNSGGGDSMAAQAQRTETEFNLPAYVTTSGAVSRSHSPADPAS-NIGLEBDI 874
 QY 2247 RKALMGKYDQWE-----SPPLSANAFNPLNASASLPPAMPITPAADGRSDHTLTSPGGG 2400
 Db 875 RKALMGSFDDKVEDHGVVMSQPMGV---VGTANTSV-----VTSGTRREBQDPSHSG 926
 QY 2401 G--KAVSGRPSRSKAKSPAP--GLASGDRPPSVSVSHSEGCNRRTPLTNRVWEDRPS 2456
 Db 927 GVCKPKLISKNSRKSPIPGQYGLTERPSVSVSHSEGDYHROT--GWAMEDRPS 984
 QY 2457 AGSTPPYNPILMLQGVMAWASPPPLPAGSG-PLAGPH---HAMDEPKPLCSQYET 2512
 Db 985 TGSQPPYNPILTKM--LSSTPPTIACAPSAVNAQAPRHOQNRKWEREPALPSAQYET 1041
 QY 2513 LSDSE 2517
 Db 1042 LSDSD 1046

RESULT 4

T13564
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
 N/Alternate names: hypothetical protein EG:49B4.1
 C/Species: Drosophila melanogaster
 C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C/Accession: T13564
 R/Spanos, L.; Papagianakis, G.; Siden-Kiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A/Description: Sequencing the distal X chromosome of Drosophila melanogaster.

Db 17 QSRYPHSVQYTFPEARHQOEFAVPDYSSHLEVSQAQSLQQOQQOQLRRRPSILSEPH 76
 QY 74 PCNERSOELHBPESHVYLPBELGKSEMERIESKRRLLELPPPLR-----PSPLLAQ 128
 Db 77 FOSDPRQE--RRSGYEQFHPGPSVVDHDSLESKRRLQVSDSHFORISAVALPLVPH-L 133
 QY 129 PAG---SBDLTQRSILTKLE-PVSPSPPHTDPELELVPRLSKEBLIQMDRDRREIT 184
 Db 134 PEGGLASSANAKDPAPGVYHAPSPPLSGQPCGDDQNASPSLSKEBLIQSDRDRRELA 193
 QY 185 MVEQOISLKLKKQOOLEEBAKPPPEKPVSPPLIESHRSIQIITYENRKKAPAHRI 244
 Db 194 KYEQOILTKLKKQOOLEEBAKPPPEKPVSPPEQKHRSIQIITYENRKKAEAKHI 253
 QY 245 LEGLQFOYELPLXNPOSDTROHENIKINOAMRKLIIYFKRBNARAKOMKPCORDO 304
 Db 254 FEGGLQPKLEPLXNPOSDTKYHENIKINOAMRKLIIYFKRBNARAKOREKLCORQO 313
 QY 305 LMEALEKKEVERIENNPBRRAKESKVRREYEQFPEIRKORELOERMQSRVQSGLSMS 364
 Db 314 LMEAEKKEVDRIENNPBRRAKESKTRREYEQFPEIRKOREQERFQ-RVQGRAGLSAT 372
 QY 365 AARSHEHSEIIDGSEONLEKOROLAVIPPMUYDQOQRIKTINGLMADPMKYK 424
 Db 373 IARSHHEISEIIDGSEONNEKOROLSVIPPMFADQORVKFINNGLMEDPMKYK 432
 QY 425 DROVANMMSBOBKETFEKFMQHPKNGFIASFLERKTVAEQVLYYLTKKNENYKSLVR 484
 Db 433 DROFANVWTDHEKEFKDKFIQHPKNGFIASFLERKSVPCVLYYLTKKNENYKALVR 492
 QY 485 RSY-RRRKSQOQOQOQOQOQOQOQPMRPSOQEKQEKKEKAEK--EERKEVEND 541
 Db 493 RNYGRKGRNQIARPSQEEKVEEK-----BEDKAEKTEKKEEKKODEBKDEDS 544
 QY 542 KEDLLKETIDTDSGENDKEKAVASKGRKTANSQGRKGRITRSMANF--ANSEBAITP 598
 Db 545 KETTEKORTAETALBEPEERQVTPRGKRTANSQGRKGRVTRMSTSAANAANAATE 604
 QY 599 QOSAEI-----ASMELNBSRWTEEMETAKGGLLEHGRNWSAIARWVSGKTVSCKNF 652
 Db 605 EPPPLPPEPPEPPISTEBVETSRKTEBEMEVAKGLVHEGRWMAIAKAVGKSKSAQCKNF 664
 QY 653 YFNFKRONLDEIIOQHKLMKEKERNARKKKAPAAASEAPPPVEDEMEBASVSG 712
 Db 665 YFNFKRHNLDNLQOQKQKASRRPREERDVSCESVASTVA---QEDIEDIAS---- 716
 QY 713 NEEEMVEEAELHNSGNVPRGSGSPATUNSSDTEISPPH-TEAK--DTGQNGPK 768
 Db 717 NEEENPEDESG-----AENSSDTEBAPSPSVEAKKSSDESENA-- 756
 QY 769 PPAITGADGPPPPPTPPRTSRAPIETTPASEATGATPPAPAPSPAPPVVPEKEKE 828
 Db 757 -----ASRGNTPEVALLEAITTPAPCASF--SSAVTTTPARRES 794
 QY 829 BE-----TAAAPVYEBEGEQ-----KPPAABELAVDTGKABEYKSECTEABE 871
 Db 795 VEAQVTDASASATEPMDVDHBEBCAGSSSVLDPAPATK--ADVNDENQV-PENTASKG 851
 QY 872 EGPAKQDAEAAEATARGALKAKEKGGSGRATTAKSSGAPO---DSDSSATCADVDE 928
 Db 852 EGDAKKEDLEST-----SEKTEARDEVDVVAEQIERPEOSDDSDSSATCSADGVD 902
 QY 929 AEGGDKNRL--LSRPSLITPTGDPANASPOKP--LDLKOLKORAAAIPI----- 976
 Db 903 GE-PERQRFVPMDAKPSLITPPGSIIL-SSPIKNNLIDLQIQRAAVIPWMSCTPCNI 960
 QY 977 -----QVTKVHEPPPEDAAPTKAPAPAPPPONLQPSDAPQOQSSSGRGK 1022
 Db 961 PIGTPVSGVALYQRIHIAHESALIE-----EQRORQEQVDLDCRSSTSPCST 1008
 QY 1023 SRSPAPADKFAAEEAQKLPQDPPCWTSGLPFPVPPPREVITKASBNHAPDSAPSVYAPGH 1082
 Db 1009 SKSP-----NRE-----W-----EVLQAPVH-----QVITNLPEGV 1034

QY 1083 PLPLGHDTAPVULPRPPTISNPPPLISSAKHPSVLERQIAISQ-----MSVOLHVPY 1137
 Db 1035 RLP-----TTPPTPPPLISSKTTVAASEK-PSFI--MGASISQGTPTGLYSHNOAYP 1086
 QY 1138 SEHAKAPVQVPMGILPLPMDPKLAPFSGVQOQOLSPRGQAGPPESLQVPTAQAESVYRG 1197
 Db 1087 QEAKRPSVGSISLQIPROESTKAAPLYITIQOEFSPRSONSQFBGLLV-RAQHEGVYRG 1145
 QY 1198 TALQSVPEGSIITKGI PSTRPVSDAITYRGSITMGTPA-----DVLVYKGTTRI-IG 1248
 Db 1146 TA-GAVOEGSITRGTTPAKSISVETISLRGSIITQGTPLPQAGIPTALVGVPSRPMIE 1204
 QY 1249 EDSPLRDRGRDELPLGCHVYIEKKGHVUSYEGEMSTQOSKDKDSSSPPHETAAPK 1308
 Db 1205 ESSPEKV--NEBASKGHVIYEGSKSHISLYDNIKNA---REGTRSPHTAHBMSI-K 1255
 QY 1309 RTYDMEGRVGRALS-----SASIEGLMGRAIPPERHSPH-HLKEOHNIRGSIITQGIPIR 1361
 Db 1256 RSYEAVESGSIKQGMWMBSPVASAPLBGLICALP--RGSPPSDLKERTVLSGSIQGTPIR 1313
 QY 1362 SYVAQEDYLARRAKLKRBSGTPPPPPSRDLTAAYTQALGRLKTPAHEGLVATYKEA 1421
 Db 1314 ATASFPEDGL-KYEQKIKRES-----PIRAFEGAI-----TKGRP-YDG-ITTIKEM 1358
 QY 1422 GRSIHEIIPREEL-----RHPELPLARPLKEGSIITQGTPLKYDTGASTTSGKKHDVRSI 1476
 Db 1359 GRSIHEIIPRODILQOESKTIPEVQSTRPIEGSISQGTPIKPDNN-SGOSAIKHNVKSL 1417
 QY 1477 IGSPGRTPPPVPHLDVAD--ARALERACYE-----ESLKSPPGASSSGSIAAGAVIV 1530
 Db 1418 ITGSHKL--PRGMLEIIPENIKVVERKTYEDVAKGEVVARHNSVVSQSVLRST---L 1472
 QY 1531 PELCKPQSPPLTYEDHQA-----PFAGHLPRGSGVTYKREPTPLQBSLSSSK-ASQDRK 1584
 Db 1473 HEAPKAQLSPQLYDSSARRTPVSYQNTISRGSPMMNR-----TSDVSSSKASHERK 1525
 QY 1585 LSTPRE-----IAKSPSTVPEHNHPHISPYEHLDRVSGVDLYRSHIPLAFDPTSLPRG 1640
 Db 1526 STLTPTPRESIPAKSPVPGVDPIVSH--SPDPHNRSSAAGEVRSRLPHLDD-ANPEH 1582
 QY 1641 IPIDAAAAYYLPRHAPNPPTYPHLYPPYLIRGYVDTAALEN-ROTIINDYITSGOMHNT 1699
 Db 1583 RALDPAAYLILQRLSTPGPSQYQY-----AMENTQTLINDYITSGOMQVNL 1633
 QY 1700 ATAMARADMLRGLSPRESSLANYAAGPRGIIDLSQVPHLPVLYPPTPGTPATAMRLA 1759
 Db 1634 -----RPDVTRGISPREQPLGLPYRA-TRGIIDLTPMP-TILVPHAGGTSTPPMDRIT 1685
 QY 1760 YLPTAPQPFSSR-HSSPFLSPGGPTHLTPTTSSSERDRDRERDREREKSIILST 1818
 Db 1686 YIPQVQYTPPRPRYNMASLSPGHPTHL--AAAASAREHEREREREKEREHERERER 1742
 QY 1819 TYVEHAP--TWPRGTQSSGSSGSGGSSRPASSHANQHSIPRTQDA-LQOR 1874
 Db 1743 ERTAAAPADLYLPRGSEQ-----PCRPCHSHGVYVPSP-SVTOETILQOR 1787
 QY 1875 PSYLHNQMGKIIITAYBPSKPTVLRSTSTSPV-----RPAATPPRATHCILGTLQVYP 1930
 Db 1788 PSVFOGTNGTSVITPDLPTAQLRIMPLPSGSPISQGLPRSRNYTADA-LAALVYMAAS 1846
 QY 1931 TLMEPVLLPKEAR-----VAPERPRAADTGAFAKPPASGELPAPASPSPK 1977
 Db 1847 APQMDVSKTESGHEAARLEENLRSSAASVBOQOLEQKULBYEKSVQCVCTSSALPSG 1906
 QY 1978 GSEPRPIY-----PVSGHATIAATPAKN--LAPHNASPDPPAPASASDPHRE 2024
 Db 1907 KAOPHASVUYSEAKQKCPPKSRUYEEELRTRKTTITANFIDVILITQIADKQARER 1966
 QY 2025 KIQSKPFSIQELELRLSGHGSYSPEGVPPVSPSLTHQKGLPKHLELDKSHLGG 2084
 Db 1967 GSQSSDSS-----SSLSSHYRTASDAIEVISPASSPAPPOKPAOYOPDMVKANQAEIN 2020

ALIGNMENTS

RESULT 1

S60255

transcription co-repressor SMRT - human

C:Species: Homo sapiens (man)

C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 05-Nov-1999

C/Accession: S60255

R:Chen, J.D.; Evans, R.M.

Nature 377, 454-457, 1995

A:Title: A transcriptional co-repressor that interacts with nuclear hormone receptors.

A:Reference number: S60255; MUID:96008552; PMID:7566127

A/Accession: S60255

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1495 <CHE>

A/Cross-references: EMBL:U57146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match 59.6%; Score 7871; DB 2; Length 1495;

Best Local Similarity 100.0%; Pred. No. 1,46-269;

Matches 1487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1031 DKFAFAEAQKLPDPCMTSGLPFPVPPREVAKSPHAPDPSAFSYPGCHPLPLGLMD 1090

DB 9 DKAFPAEAQKLPDPCMTSGLPFPVPPREVAKSPHAPDPSAFSYPGCHPLPLGLMD 68

QY 1091 TAPVLPVLPPTISNPPLISSAKHPSVLERQIGAIISQGMVQLHVPYSEHAKAPVGVPM 1150

DB 69 TAPVLPVLPPTISNPPLISSAKHPSVLERQIGAIISQGMVQLHVPYSEHAKAPVGVPM 128

QY 1151 GLPLMDPKLAPSSGKQKOLSPRGQAPRESIGVTAQEAVALRGTAIGSVPGSITK 1210

DB 129 GLPLMDPKLAPSSGKQKOLSPRGQAPRESIGVTAQEAVALRGTAIGSVPGSITK 188

QY 1211 GIPSTRPDSATYRGSITHTGTPADLYKGTITRIIGEDSPRLDGRSDLPKGVIVY 1270

DB 189 GIPSTRPDSATYRGSITHTGTPADLYKGTITRIIGEDSPRLDGRSDLPKGVIVY 248

QY 1271 EGKKGVALSYEGGMSVTCSEKEDRSSGPHETAAFKRTYDMMEGVRAISSASIEGL 1330

DB 249 EGKKGVALSYEGGMSVTCSEKEDRSSGPHETAAFKRTYDMMEGVRAISSASIEGL 308

QY 1331 MGRAIPERHSPIHLKQOHIRGSIITGIPRSYEAEDVIRRAKLLKRGTPPPPPS 1390

DB 309 MGRAIPERHSPIHLKQOHIRGSIITGIPRSYEAEDVIRRAKLLKRGTPPPPPS 368

QY 1391 RDLTEAKTQALGPLKPAHEGLVATVYKAGRSIHETPREELHTPELPLAPPLKEGS 1450

DB 369 RDLTEAKTQALGPLKPAHEGLVATVYKAGRSIHETPREELHTPELPLAPPLKEGS 428

QY 1451 ITGOTPLKYDTGASTGSKGDVRSLLIGSPGRTFPVPHPLDVMADALERAQYESSLSKS 1510

DB 429 ITGOTPLKYDTGASTGSKGDVRSLLIGSPGRTFPVPHPLDVMADALERAQYESSLSKS 488

QY 1511 RPTGASSGSIARGAVIYBELGKRPQSLTYEDHGAFFAGHILPRSSPTMREPTRLQ 1570

DB 489 RPTGASSGSIARGAVIYBELGKRPQSLTYEDHGAFFAGHILPRSSPTMREPTRLQ 548

QY 1571 EGSLSSSKASQDRKLTSTPREIAKSPHSTVEHHPHIPISPEHLIRGSGVDLRSRSHPL 1630

DB 549 EGSLSSSKASQDRKLTSTPREIAKSPHSTVEHHPHIPISPEHLIRGSGVDLRSRSHPL 608

QY 1631 AFDPTSIPIRGIPLDAAAAYVLPRLHAFNPYVPHLYRGPYDTALENRQTIINDYI 1690

DB 609 AFDPTSIPIRGIPLDAAAAYVLPRLHAFNPYVPHLYRGPYDTALENRQTIINDYI 668

QY 1691 TSOQMHNTATAMORADMTKGLSPRESSALNVAAPRGITDLSQVPHLPVYVPPRG 1750

DB 669 TSOQMHNTATAMORADMTKGLSPRESSALNVAAPRGITDLSQVPHLPVYVPPRG 728

QY 1751 PATAMRLAVALPTAPQPFSSRHSSPLSPGPTALTKPTTSSERRDRDRDRDR 1810

DB 729 PATAMRLAVALPTAPQPFSSRHSSPLSPGPTALTKPTTSSERRDRDRDRDR 788

QY 1811 EKSLITSTTVERAPITWRPCTEQQSSGSSGGGSSSRPASHSHAHQSPISPRQDA 1870

DB 789 EKSLITSTTVERAPITWRPCTEQQSSGSSGGGSSSRPASHSHAHQSPISPRQDA 848

QY 1871 LQORPSVLAHTGKGGITTAVERPSKPTLARSSTSSPPVPAATPPATPCPLGGLDGV 1930

DB 849 LQORPSVLAHTGKGGITTAVERPSKPTLARSSTSSPPVPAATPPATPCPLGGLDGV 908

QY 1931 TIMEPVLPKEADRVARPERPRADTGHAFLAKPPARSGLEPASSPSKSGSEPRPLVPV 1990

DB 909 TIMEPVLPKEADRVARPERPRADTGHAFLAKPPARSGLEPASSPSKSGSEPRPLVPV 968

QY 1991 HATITARPANLAPHHASPPPPAPASAPPHREKTQSKFPIQELERLRLGHSSTYSP 2050

DB 969 HATITARPANLAPHHASPPPPAPASAPPHREKTQSKFPIQELERLRLGHSSTYSP 1028

QY 2051 EGVEPVSPVSPSLTHDKGLPKLLEELDKSHLGEELRPKOPGVKLGGEAAHLPHLRPLP 2110

DB 1029 EGVEPVSPVSPSLTHDKGLPKLLEELDKSHLGEELRPKOPGVKLGGEAAHLPHLRPLP 1088

QY 2111 ESQPSSEPLIQTAPGVKQHVYTLAQHISEVITQDYTRHHPOQLSAPLAPLYSPGAS 2170

DB 1089 ESQPSSEPLIQTAPGVKQHVYTLAQHISEVITQDYTRHHPOQLSAPLAPLYSPGAS 1148

QY 2171 CPVLDLRPPSDLYLPPPDHGAAPKSGPHSEGGKRSSEPKTSTVLGGEGGIEVSPDEG 2220

DB 1149 CPVLDLRPPSDLYLPPPDHGAAPKSGPHSEGGKRSSEPKTSTVLGGEGGIEVSPDEG 1208

QY 2231 MTEPGHRSAYVPLLYNDGQTEPSRMGSKSPGNTSPPAFPSKLTSSNAMYKSKQOI 2290

DB 1209 MTEPGHRSAYVPLLYNDGQTEPSRMGSKSPGNTSPPAFPSKLTSSNAMYKSKQOI 1268

QY 2291 NKLANTHREPEYNIISQPTETIFNMPAITGTGLMTYRQAVQEHASTNGLAIIKRL 2350

DB 1269 NKLANTHREPEYNIISQPTETIFNMPAITGTGLMTYRQAVQEHASTNGLAIIKRL 1328

QY 2351 MGKYDQWESPPISANAFNPLNASSILPAMPITTAADGRDHTLTSGGGGKAVSGRPS 2410

DB 1329 MGKYDQWESPPISANAFNPLNASSILPAMPITTAADGRDHTLTSGGGGKAVSGRPS 1388

QY 2411 SRKAKSPAPGLASGDRPVSYSVHSEDCNRRTPLTNRVEDRPSAGSTPFPPNPLMR 2470

DB 1389 SRKAKSPAPGLASGDRPVSYSVHSEDCNRRTPLTNRVEDRPSAGSTPFPPNPLMR 1448

QY 2471 LQAGVNASPPPPGLPAGSGPLAGPHHAMDEPRYLCSQYETLSDSE 2517

DB 1449 LQAGVNASPPPPGLPAGSGPLAGPHHAMDEPRYLCSQYETLSDSE 1495

RESULT 2

S60254

nuclear receptor co-repressor N-COR - mouse

C:Species: Mus musculus (house mouse)

C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 08-Oct-1999

C/Accession: S60254

R:Heinzel, A.M.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ryan, A.;

Nature 377, 397-404, 1995

A:Title: Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear

A:Reference number: S60254; MUID:96008539; PMID:7566114

A/Accession: S60254

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-2453 <HOE>

A/Cross-references: EMBL:U5312; NID:g1022717; PIDN:AA817125.1; PID:g1022718

Query Match 31.7%; Score 4187; DB 2; Length 2453;

Best Local Similarity 41.0%; Pred. No. 7,76-140;

Matches 1103; Conservative 346; Mismatches 796; Indels 444; Gaps 106;

QY 16 EPRYPHSLSYVQIARTHTDVLGLLEYO--HHSRDVASHLSPGSITQPORRPSLSSEFQ 73

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 21:53:54 ; Search time 60 Seconds
(without alignments)
4035.235 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQTRATEPRYP.....WDEPKPLCSQYETLDSR 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7871	59.6	1495	2	transcription co-r
2	4187	31.7	2453	2	nuclear receptor c
3	1272	9.6	1047	2	hypothetical prote
4	574	4.3	5327	2	microtubule-associ
5	552	4.2	2649	2	hypothetical prote
6	526	4.0	2142	2	MHC class III hist
7	520	3.9	2187	2	nascent polypeptid
8	512	3.9	5762	2	proline-rich pepti
9	508.5	3.8	1872	2	MHC class III hist
10	504	3.8	3942	2	Bassoon protein -
11	503.5	3.8	7962	2	elastic titin - hu
12	503	3.8	1870	2	MHC class III hist
13	502.5	3.8	2715	2	eyelid - fruit fly
14	498	3.8	3938	2	Bassoon protein -
15	492.5	3.7	2274	2	microtubule-associ
16	477	3.6	2282	2	DNA-binding protei
17	470.5	3.6	4957	2	ADP-ribosylating
18	470.5	3.6	5262	2	hypothetical prote
19	445.5	3.4	1952	2	hypothetical prote
20	436	3.3	1151	2	high molecular mas
21	424.5	3.2	2464	1	microtubule-associ
22	419	3.2	5170	2	hypothetical prote
23	416.5	3.2	1791	2	hypothetical prote
24	412.5	3.1	1226	2	hypothetical prote
25	407.5	3.1	1464	1	collagen alpha 1(I
26	406.5	3.1	2364	2	microtubule-associ
27	404	3.1	1274	2	hypothetical prote
28	403	3.0	1110	2	NF-160 - sea lamp
29	397	3.0	1670	1	collagen alpha 3(I

30	395	3.0	1520	2	T00273	hypothetical prote
31	394	3.0	2783	1	A41948	alpha-fetoprotein
32	391	3.0	3924	2	S37431	ankyrin 2, neurona
33	389.5	2.9	1188	2	S49915	extensin-like prot
34	389.5	2.9	3869	2	A48205	All-1 protein +GTE
35	389	2.9	1320	2	JC5610	TCOF1 protein - mo
36	383.5	2.9	3968	2	A44265	trithorax homolog
37	383	2.9	2944	2	A54849	collagen alpha 1(V
38	377	2.9	1691	1	S22917	collagen alpha 5(I
39	375	2.8	1006	2	T42721	atrophin-1 related
40	374.5	2.8	2722	2	T20532	hypothetical prote
41	373.5	2.8	1184	2	S50832	atrophin-1 - human
42	373.5	2.8	1494	2	T14355	protein-tyrosine-p
43	373.5	2.8	2938	2	T30249	cell proliferation
44	373	2.8	1763	2	S16356	collagen alpha 2(I
45	372.5	2.8	1173	2	T31421	C-terminal domain-
46	372.5	2.8	1184	2	G01763	atrophin-1 - human
47	368.5	2.8	2738	2	E88320	protein F07A11.6 I
48	367	2.8	1611	2	T38236	hypothetical prote
49	360	2.7	2232	2	T34434	hypothetical prote
50	358	2.7	1618	2	S21424	nestin - human
51	356.5	2.7	2526	2	T20531	hypothetical prote
52	355	2.7	2090	2	S26058	probable transform
53	354	2.7	3507	2	T34513	hypothetical prote
54	350.5	2.7	2843	1	RBHUP	adenomatous poly
55	349.5	2.6	1840	2	T30250	GM1 protein - mous
56	348	2.6	1669	1	CGMS4B	collagen alpha 1(I
57	348	2.6	3256	2	A48666	cell proliferation
58	346.5	2.6	1018	2	S44758	CI4B9.6 protein -
59	345	2.6	1843	2	S18803	collagen alpha 1(V
60	344.5	2.6	1546	1	CGHUE	collagen alpha 2(X
61	344	2.6	2845	2	I49505	adenomatous poly
62	343.5	2.6	1464	1	S59856	collagen alpha 1(I
63	343	2.6	1466	1	CGHUTL	collagen alpha 1(I
64	342.5	2.6	1020	1	QPHDH	neurofilament trip
65	342	2.6	5105	2	T32650	hypothetical prote
66	341	2.6	1200	2	A46194	neurofilament prot
67	341	2.6	1453	2	S21626	collagen alpha 1(I
68	341	2.6	1571	2	T00062	hypothetical prote
69	339.5	2.6	1690	1	CGHUIB	collagen alpha 4(I
70	338.5	2.6	2897	2	B48666	cell proliferation
71	338	2.6	13055	2	T16580	hypothetical prote
72	337	2.6	1492	2	A40333	collagen alpha 1(I
73	337	2.6	1634	2	T26517	hypothetical prote
74	336	2.5	839	2	F75518	hypothetical prote
75	335.5	2.5	3570	2	T45025	myosin I precursor,
76	335	2.5	1344	1	A35175	myosin I precursor,
77	335	2.5	1418	1	T45467	collagen alpha 1(I
78	335	2.5	1460	1	EDBRIF	immediate-early pr
79	333.5	2.5	1487	1	CGHUC6	collagen alpha 1(I
80	331	2.5	1838	1	CGHUIV	collagen alpha 1(I
81	330	2.5	1367	1	S48478	glucan 1,4-alpha-g
82	329.5	2.5	916	2	A27864	neurofilament trip
83	329.5	2.5	1819	2	T32008	hypothetical prote
84	329	2.5	1280	1	T00365	hypothetical prote
85	326.5	2.5	1414	1	S23809	collagen alpha 2(I
86	326.5	2.5	1469	2	T09212	basal transcrip
87	325	2.5	1829	2	T26135	hypothetical prote
88	324.5	2.5	2022	2	T48818	hypothetical prote
89	324	2.5	3149	1	QOBER	glucan 1,4-alpha-g
90	323.5	2.4	1497	2	I49607	BPLF1 protein - hu
91	323.5	2.4	1603	2	S23810	procollagen type V
92	323.5	2.4	4549	2	T20771	collagen alpha 1(X
93	323.5	2.4	4667	2	T20774	hypothetical prote
94	323	2.4	1446	1	A45344	hypothetical prote
95	322.5	2.4	1684	2	JW0057	immediate-early pr
96	320.5	2.4	1669	1	CGHUIB	collagen alpha 1(I
97	319	2.4	2361	2	T25752	hypothetical prote
98	318.5	2.4	1094	2	S49313	protein kinase - s
99	318	2.4	1924	2	S27923	gene Lf3 protein -
100	318	2.4	1467	2	B41182	collagen alpha 1(I